

Time for Japan to shine?

Talks last weekend on choosing a site for ITER, the fusion project, ended in stalemate. But ITER deserves to proceed, and Japan's commitment to it is strongest.

Fusion power is sometimes mocked for its alleged unfeasibility — it's been 40 years away since about 1955, detractors say. But for those who are optimistic about man's ability to meet tall technical challenges, its allure will never fade.

The power of the Sun and the stars derives from the fusion of hydrogen nuclei into helium at temperatures of about 100 million kelvin. Scientists and engineers already know a fair bit about how to harness this process here on Earth. The progress of existing, small fusion experiments has been steadily impressive.

Converting that progress into something akin to a fusion power plant will be difficult, but not impossible. The fuel of such a plant will instantly evaporate its surroundings unless it is suspended, probably by a magnetic field. Fusion itself produces an avalanche of highly destructive neutrons, which will embrittle any surrounding structures, requiring constant replacement.

For the faint-hearted, there are far more reasonable solutions available to our impending energy needs. Fission power can come back — preferably in someone else's backyard. Coal can be burned — there's a lot of it, probably enough to choke the planet in greenhouse gases. Windmills can be built, even if they will prove a blot on the beautiful landscape of the Scottish Highlands. And peasants in China and Africa can tend to their own energy needs by burning organic manure in biodegradable ovens.

None of these energy solutions should be glibly dismissed, of course. But neither should the immense potential of fusion. There are several approaches to harnessing it: some use inertial confinement, which applies immense heat and pressure to a tiny fuel capsule, producing a controlled explosion.

But magnetic confinement, in which the plasma fuel is held aloft in a magnetic field, is the best-developed avenue. This would probably be done in a doughnut-shaped vessel called a tokamak, invented in Russia. Successful tokamak experiments have been built in Japan, the United States, Britain, Russia and elsewhere. But for years, plasma physicists have sought a larger machine that would hold burning plasma for a sustained period, so they can study its behaviour in detail. Twenty years ago at a summit meeting in Reykjavik, Ronald Reagan and Mikhail Gorbachev bought the idea, and ITER — then the International Thermonuclear Experimental Reactor — was born.

Political problems

It hasn't exactly been plain sailing since then. The energy crisis has faded from the public mind, although it is now edging back. Fusion has suffered collateral damage from public unease with nuclear power — although fusion would produce no high-level radioactive waste and much less lower-level waste than fission. ITER's design phase was nonetheless completed in 1998, only for Congress to pull the plug on US participation in construction. That encouraged the remaining partners — Russia, Japan and Europe — to shrink the project. Now the United States has rejoined it, and China and South Korea are in too.

The time has come to choose a site. The two contenders are the European Union (EU), backing a site in Cadarache, France, and Japan, which would build ITER at Rokkasho. In a wretched and unnecessary

turn of events, site selection has become heavily politicized. The rot set in when Spencer Abraham, the US energy secretary, took the unprecedented step last year of butting into Europe's choice of its candidate site, expressing a preference for an earlier bid by Spain, presumably to smite France for its stance on the invasion of Iraq. But France won the European contest. On Saturday, Europe, Russia and China backed the Cadarache site while Japan, the United States and South Korea supported Rokkasho.

ITER is easily the largest international scientific collaboration on the table at present and it should not be bogged down in geopolitics. But the deadlock of recent weeks can be broken.

The way forward

A solution will involve providing the loser with a prominent role in a broader nuclear-fusion programme. This is much less of a problem than it might seem. International participants need a remote control facility to widen the use of the experiment. More importantly, a separate materials test centre built around a neutron source has always been required, in advance of the engineering prototype reactor, known as Demo, that will follow ITER. Such a programme would not only offer a compromise by giving both Europe and Japan a major fusion facility, but it would be a major boost to fusion science itself and to making better use of ITER.

France has considerable technical know-how in this field and a record of excellence in nuclear engineering. Cadarache is a superbly qualified site. But despite the rallying of EU states around France's bid, political and public support in Europe are less than whole-hearted.

Japan's bid is for a green-field site in the remote north of the country, and its selection would raise some logistical challenges and local opposition. But Japan's formidable state apparatus is now firmly behind the project. With no indigenous energy resources to speak of, the country's commitment to long-term energy sources is greater than Europe's, and much greater than that of the United States. The ITER negotiations have had a high public profile in Japan, and the nation genuinely yearns for an international scientific project of this calibre. Its existing JT-60 tokamak is on a par with any other facility in the world, and Japan's political structure and engineering prowess would be conducive to ITER's successful completion.

Yearning is no substitute for a capability to deliver, however, and last weekend's meeting decided that another gathering should be held in Vienna next month, where experts will be asked to go through a full comparative assessment of the sites as they stand, point by point. If Europe's case is technically strongest, then Japan's compensation should include international contributions to an upgrade to JT-60 to achieve critical science on the way to Demo. An upgraded JT-60 would, in particular, like ITER, be capable of sustained fusion burns, and complement ITER by making possible the comparison of plasma behaviour on different scales.

But if the technical strengths are well balanced, Europe should turn the other cheek to the Bush administration's mischief-making and break the impasse, by negotiating a deal to build computing and materials research centres in France and Spain, and ITER itself in Japan, the nation that most needs it and most ardently supports it. ■





Poor cows
US edges towards wider testing of cattle for BSE
p766



On the alert
Initiative presses labs to guard against bioterrorism
p767



POP chart
Countdown begins for persistent organic pollutants
p768



Brave new world
US army joins games company to create training simulation
p770

Media attack prompts editorial backlash against MMR study

Jim Giles, London

A leading medical journal has taken the unusual step of distancing itself from one of its own papers and attacking its findings.

Editors at *The Lancet* launched a pre-emptive media strike on 18 February after learning that *The Sunday Times* was investigating undeclared conflicts of interest in a 1998 paper linking autism with the measles, mumps and rubella (MMR) vaccine. The editors went public two days before the story was due to be published, declaring that the autism paper was "flawed" and should not have been published in its original form.

The paper triggered huge public controversy in Britain over the safety of the MMR vaccine. In it, lead author Andrew Wakefield, a gastroenterologist then at the Royal Free Hospital in London, described a group of eight children with behavioural problems and noted that their parents reported problems beginning within two weeks of the children receiving MMR jabs (A. Wakefield *et al.* *The Lancet* 351, 637–641; 1998).

At a press conference before publication, Wakefield said that the three vaccines should not be administered in combination until the issue had been resolved. Subsequent epidemiological studies failed to find a link with autism and led many of Wakefield's co-authors to distance themselves from his claims. But this did not stop the paper and Wakefield's continuing criticism of the vaccine from having a dramatic effect.

Conflicting, high-profile television and newspaper reports about the safety of the triple vaccine confused parents. Prime Minister Tony Blair, who urged parents to have their children vaccinated, came under pressure in parliament to reveal whether his three-year-old son had received the jab. National take-up of the MMR jab dropped from around 90% in 1998 to its current level of less than 80%.

But unknown to *The Lancet*, Wakefield had been asked to conduct a separate study into whether parents whose children had allegedly been harmed could sue manufacturers of the MMR vaccine. The Legal Aid Board, which then administered government funding for such claims, paid Wake-



Tough call: UK parents panicked following claims that the MMR vaccine might cause autism.

field £55,000 (US\$103,000) towards the costs of that study, according to *The Sunday Times*. The newspaper said that at least four children took part in both studies.

In last week's statement, editors at the journal said Wakefield's failure to disclose the existence of the legal-aid study was a clear breach of its conflict-of-interest policy, which in 1998 stated that authors should reveal anything "that would embarrass you if it were to emerge after publication and you had not declared it".

Astrid James, the journal's deputy editor, says that the actual data in the paper are not being questioned, but that reviewers and editors would have viewed Wakefield's interpretation differently had they known the source of the funding. "The link with MMR is flawed," says James. "We would have asked authors to change that section." The journal will publish a commentary on the paper, together with statements from the original authors, in the next few weeks.

The Lancet's verdict is in line with mainstream opinion on the Wakefield paper, but the journal's reaction has surprised some people in scientific publishing. Richard

Smith, editor of the *British Medical Journal* (*BMJ*), agrees that the paper is flawed on scientific grounds, but questions whether a declaration of conflict of interest would have affected the decision to publish.

Smith says that the *BMJ* regularly publishes papers by authors who have been paid to act as expert witnesses in court cases. The journal also accepts papers from researchers working for pharmaceutical companies. "Drug company employees have more extreme conflicts of interest," Smith says, adding that journals usually deal with these by disclosing them — not by asking authors to change their paper.

Wakefield has left the Royal Free and is now director of an autism research centre in Florida. He accepts that he was funded by the Legal Aid Board, but says that the results of the *Lancet* paper did not affect the possibility of a case being brought, and his funding was therefore irrelevant.

He could now face an investigation by the General Medical Council, the body that regulates doctors — a prospect which Wakefield said in a statement he welcomed as an opportunity to clear his name. ■

S. ASHFELD/SPL

Poor communication blamed for delays to US visas

Geoff Brumfiel, Washington

Foreign students and scientists coming to the United States are suffering big delays caused by immigration restrictions, a congressional report has confirmed.

The report, which is set for release this week, tracked 71 science students and scholars who underwent security screenings in the wake of the 11 September 2001 terrorist attacks. It found that they waited a mean period of 67 days for their visas — well above the state department's 30-day target.

"The report gives weight to what a lot of us were guessing at from anecdotal evidence," says Wendy White, who works on international issues at the National Academy of Sciences.

The House Committee on Science commissioned the study last year from the Government Accounting Office (GAO), after hearings showed that foreign scientists were facing lengthy visa delays (see *Nature* 422, 457; 2003).

The delays are particularly bad for scientists from countries such as India, China and Russia, who must undergo frequent interagency security reviews that involve the FBI, the CIA and the Department of Homeland Security.

The GAO says that most delays are caused by communication lapses, either between these departments or within the state department, which administers the embassies and consulates. It can take up to two-and-a-half months for a request to get from an embassy to the FBI, and as long as 45 days for the FBI's clearance to reach the state department, it says.

The report also says that consular staff "lack guidance" on when to apply security reviews, and receive little feedback on whether they are providing the right information about an applicant.

Stuart Patt, a spokesman for the state department, defended its position, saying that it was hiring new consular officers and providing some scientific training for existing ones. He added that the department would soon install a new computer system, which should aid internal communication.

But White's office has received more than 1,100 complaints from foreign scientists since last June. She says that the state department's plans will help to reduce delays, but that much more remains to be done. "The mechanics of the process are getting better," she says, "but the overall feeling that the United States is no longer a welcoming place persists." ■



Still reeling: the US effort to monitor mad cow disease has yet to find its feet, according to congressmen.

Mad cow's standing puts downer on testing strategy

Erika Check, Washington

The US Department of Agriculture (USDA) is facing growing pressure from Congress to overhaul its surveillance regime for mad cow disease.

A powerful congressional committee says that the agriculture secretary, Ann Veneman, gave the public wrong information about the first mad cow found in the United States in December. Veneman said at the time that the cow had been targeted for testing because it was a 'downer' — an animal that arrived at the slaughterhouse too sick to walk (see *Nature* 427, 5; 2004).

But two employees at the slaughterhouse in Washington state where the animal was killed, and the hauler who took the cow there, have all told congressional investigators that the cow wasn't a downer at all. The revelation seems to undermine Veneman's assertion that the department's existing focus on testing obviously sick animals was vindicated by the cow's detection.

In an unusual bipartisan move, the top Republican and the senior Democrat on the House Committee on Government Reform wrote to Veneman on 18 February, urging her to resolve the issue.

According to the letter, the co-manager of the slaughterhouse faxed a USDA office on 6 January to state that the mad cow was not a downer. He said that the animal was healthy, and had been tested under a contract whereby the USDA paid the slaughterhouse \$10 per cow for test samples.

The revelations could force the agriculture department to change its current plan to test just 40,000 cows for mad cow disease

this year. "If the new information is accurate, USDA's surveillance programme may need to be significantly expanded," wrote congressmen Tom Davis (Republican, Virginia) and Henry Waxman (Democrat, California) in their letter. "The new information also raises questions about USDA's credibility."

Earlier this month, an international review team chaired by Ulrich Kihm, Switzerland's chief veterinary officer, recommended that the USDA expand its programme to test all cows over 30 months old as well as some younger, healthy animals (see *Nature* 427, 575; 2004). In their letter, the congressmen agreed: "We believe USDA should either follow the recommendations of these independent experts and expand mad cow testing substantially or provide a compelling reason for not doing so," they wrote.

And while the agriculture department seemed reluctant to accept that advice at first, its position appears to be becoming more flexible. "Everything is in flux right now," says Jim Rogers, a spokesman for the USDA's Animal and Plant Health Inspection Service. "Last I heard we were on target for testing 40,000 animals a year, but that could change, for sure."

Rogers says that the agriculture department is also trying to figure out how to find and test downed animals, now that they are excluded from the food supply and will no longer be tested at slaughterhouses. He added that the department's inspector general would investigate the allegations that the first mad cow was not actually a downer. ■

Labs urged to pre-empt bioterrorism threat

Erika Check, Washington

An initiative is under way to get biotechnology companies to reduce the risks that their staff or labs could nurture bioterrorism — either by accident or by design.

Two non-profit groups, the International Institute for Strategic Studies (IISS) and the Chemical and Biological Arms Control Institute, announced the initiative in Washington on 20 February.

They are hoping that pharmaceutical and biotechnology companies will voluntarily join a scheme to prevent the misuse of tools at their researchers' disposal.

Michael Moodie, president of the arms control institute, and IISS president Terence Taylor are drawing up a charter for companies to sign certifying that they have taken steps to address bioterrorism risks. They also plan to bring industry leaders, academic scientists and government officials together at forums to discuss the issue.

Taylor says that the initiative is inspired by a report released last October by a National Academy of Sciences panel. The panel, chaired by Gerald Fink, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, recommended steps to prevent the misuse of biological knowledge (see *Nature* 425, 647; 2003). The Bush administration initially said that it would respond to the report within a month, but hasn't yet done so.

Government officials welcomed Taylor and Moodie's plan, saying that they have few



Risk of abuse? A large-scale fermenter used in production of monoclonal antibodies for vaccines.

mechanisms for keeping track of drug-industry labs. "Now that the biological weapons protocol is dead, we have no interactions with those communities," says Jennie Gromoll, an official handling biological weapons issues at the state department.

Others cautioned that it might be difficult for the initiative to draw a response from the private sector. Eileen Choffnes, an arms-control expert at the national academy, says

that many companies aren't convinced that their work could contribute to bioterrorism. "Unless you can define the problem with objective evidence, many people are just going to go through the motions," she says.

Una Ryan, chief executive of Avant in Needham, Massachusetts, which is working on developing an anthrax vaccine, supports the move "because there's no agency constantly monitoring the situation".

Electron movements pinned down to the split second

Mark Peplow, London

Physicists in Austria say they that have observed events separated by the shortest time interval ever, and plan to use the technique to study atomic phenomena.

A group led by Ferenc Krausz of Vienna University of Technology used pulses of laser light to watch electrons moving around atoms, and were able to distinguish events that took place 100 attoseconds — or 10^{-16} seconds — apart (see page 817 of this issue).

Krausz and his team say that the research could improve understanding of the role of electrons in superconductivity and in giant magnetoresistance, a magnetic effect used in data-storage devices. "It's only on the attosecond timescale that you see these things happening," says Roger Falcone, a physicist at the University of California, Berkeley.

The attosecond (10^{-18} s), also known as a quintillionth of a second, is the timescale of atomic events. In Niels Bohr's 1913 model of a hydrogen atom, it takes about 150 attoseconds for an electron to orbit the proton.



Just a tick: scientists are using ever-shorter timescales to investigate chemical reactions.

Krausz used ultraviolet-light pulses of 250-attosecond duration to excite electrons in atoms of neon, raising them to a higher energy level. These electrons come out of the atoms with a range of different momenta. A second pulse of light, delivered just after the first, then gives the electrons an extra jolt. Changing the delay between these two pulses produces a different distribution of momentum values for the electrons, which

the scientists measured with a sensitive electron detector called a 'streak camera'. Comparing these distributions gave the team a snapshot of how the electrons differed when hit with an early or late second pulse.

Krausz thinks that the technique could soon be refined to track events a few tens of attoseconds apart, telling physicists how electrons rearrange themselves inside an atom when they move between different orbits.

Recent progress in using fleeting laser flashes to study short time scales has opened up the field of femtochemistry, in which the motion of atoms and molecules is tracked on the femtosecond (10^{-15} s) timescale, to see how they rearrange themselves during chemical reactions. Probing even shorter timescales could provide more insight.

"Every time we get shorter pulses, chemists say that we don't really need them," says David Klug, a photochemist at Imperial College, London. "But then they find a way to use them. Once it becomes convenient, people will find applications for it."

Treaty calls time on long-term pollutants

Jim Giles, London

The clock is ticking for a troublesome set of chemical pollutants. Last week, a three-month countdown began towards the day when the Stockholm Convention on Persistent Organic Pollutants (POPs) comes into force.

But environmental-health specialists warn that the first phase of the agreement, drawn up in 2001, represents a relatively easy win. From May, the convention will ban the use of 12 types of pollutant of limited economic importance; but subsequent constraints on more financially valuable compounds will be fiercely fought by chemical manufacturers, the specialists predict.

The countdown towards implementation was triggered on 17 February, when France became the fiftieth country to ratify the agreement. Several countries, including the

United States and Britain, have yet to ratify it, but are preparing legislation to do so.

Environmentalists have campaigned for years against POPs, a group of very stable compounds that includes dioxins and the pesticide chlordane. POPs accumulate in animals and plants at the end of the food chain, and have been linked to a range of human health problems.

Chemical manufacturers accept that the 12 substances covered by the first phase of the convention are dangerous, and have already phased some of them out. But Bo Wahlstrom, a Geneva-based scientific adviser at the United Nations Environment Programme, which brokered the treaty, predicts that it will be far harder to incorporate compounds that are still made and sold in large quantities.

Substances that are likely to spark disagreement include brominated flame retar-

dants such as hexabromocyclododecane (HBCD) and decabromodiphenyl ether (decaBDE). Tens of thousands of tonnes of these compounds are made every year, primarily as flame-retarding additives for textiles, electrical equipment and building materials. The retardants can enter the environment during manufacture and disposal, and their stability means that they can build up in humans and animals over time.

The jury is still out on whether the current level of exposure to these compounds presents a health risk to people and animals, say environmental-health researchers. But, they add, HBCD and decaBDE meet some of the criteria for being classed as POPs.

Steve Robinson, chemicals assessment manager at the Environment Agency in Wallingford, near Oxford, UK, says that HBCD is toxic and accumulates in living tissue. He adds that industry is now running tests to assess another criterion for it to be classed as a POP: whether or not it persists in the environment. So far, the only evidence for decaBDE as a possible POP is that it has been shown to build up in the tissue of animals at the end of the food chain, such as harbour seals, birds of prey and dolphins (L. S. Birnbaum and D. F. Staskal *Environ. Health Perspect.* **112**, 9–17; 2004).

The status of such contentious compounds will become clear when European Union nations complete a risk assessment of flame retardants later this summer. The Bromine Science and Environment Forum, which represents bromine manufacturers, says that the industry is working with officials on this assessment and that, so far, the studies show no need for decaBDE or HBCD to be added to the convention. ■

♦ www.pops.int



Harbour seals are among the creatures in which persistent organic pollutants may accumulate.

Closure threat propels astronomers to institute's aid

Quirin Schiermeier, Munich

Top astronomers are protesting that a proposal to close the Astronomical Institute at the University of Basle in Switzerland ignores the institute's strong track record in research.

The university's governing council recommended on 22 January that the department, along with the university's institutes of geology and Slavic studies, should close by 2008.

Rolf Soiron, the businessman who chairs the council, says that the closures were needed to consolidate finances and avoid duplication of training and research facilities in Switzerland.

The Swiss government, which supports

the consolidation of small university departments, has backed the plan. "It is a courageous and exemplary move," says Charles Kleiber, the Swiss state secretary for education and science.

But astronomers are complaining that the decision was made in response to low student numbers, without any external scientific assessment of the institute's performance. In letters to Soiron, more than 100 astronomers from around the world protested against the decision.

The institute has "produced by far the most careful and thorough dynamical models of the Milky Way that are available anywhere in the world", wrote Scott Tremaine, chair of astrophysical sciences at

Princeton University, New Jersey.

Astronomers also criticized the treatment of Eva Grebel, a rising star among galactic astronomers, who won tenure at the institute only last September.

"People will think that in Switzerland you can easily lose your job, even though you are tenured. This is a blow to our reputation," says Willy Benz, an astrophysicist at the University of Berne and a member of the Swiss Science and Technology Council.

The university defends the plan. "This was not done in a rush. The board spent months coming up with a balanced reform," says Maria Schoch, a spokeswoman for the rector. ■



Sense of urgency: scientists on the *Discovery* deploy moorings that carry sensors to the ocean floor.

Gulf Stream probed for early warnings of system failure

Quirin Schiermeier, Munich

Climate researchers set sail from the Canary Islands today to begin an ambitious, four-year programme that will assess the behaviour of currents, such as the Gulf Stream, in the north Atlantic Ocean.

The US\$20-million programme is part of a wider investigation into rapid climate change, known as Rapid. It will take the most detailed look yet at the strength, structure and variability of the currents that carry warmth northwards in the Atlantic.

Climatologists worry that global warming could disrupt these currents, which make Western Europe's climate far warmer than other parts of the world at the same latitude. Without the Gulf Stream, for example, the south of England would be as cold as Iceland.

Researchers think the currents are caused by a combination of wind, differences in water density, and the special geometry of Atlantic Ocean basins and the surrounding continents. They think the system may have broken down before, driving large, abrupt changes in climate (see *Nature* 364, 203–207; 1993).

Given the extent of global climate change at the moment, some suspect this could happen again. Models suggest that currents are already affected by increased freshwater flow from precipitation, river runoff and ice-sheet melting (see *Nature* 378, 145–149; 1995).

The Rapid team on board the British research vessel RRS *Discovery* will try to establish if this is really happening. "Complete collapse of the north Atlantic circulation is a worst-case scenario," says Stefan Rahmstorf, a climate modeller at the

Potsdam Institute of Climate Impact Research in Germany. "But no one really has any firm idea of what is going on out there, and that is why this project is so important."

The team will use 22 moorings across the subtropical Atlantic: near the Canary Islands, on the Mid-Atlantic Ridge and off the Bahamas. Sensors will travel up and down wires from buoys to the moorings on the sea floor. Differences in water density will be calculated from the temperature and salinity measured throughout these water columns. With US measurements from the Florida Strait and satellite observation of wind-driven surface currents, these will help researchers understand water flow in the north Atlantic.

"It is unlikely that we will detect dramatic changes within the next four years," says Jochem Marotzke, an oceanographer at the Max Planck Institute of Meteorology in Hamburg, and one of the principal investigators in Rapid. "But we will learn lots of exciting things about ocean circulation, and work out how to design a stable monitoring system for the next few decades." Rapid is being jointly sponsored by Britain's Natural Environment Research Council and the US National Science Foundation.

One long-term goal, says Marotzke, is an early-warning programme that would raise the alarm if the present system was close to failure. Even if nothing could be done, society could prepare for the results, he says. A German–Norwegian project called Integration is already assessing the impact of a failure on climate, fisheries and agriculture. ■

► www.soc.soton.ac.uk/rapidmoc

Bioprospectors edge towards deal with developing countries

Rex Dalton, Kuala Lumpur

Guidelines for researchers, companies, governments and universities engaged in bioprospecting are to be drawn up within two years, representatives of nearly 200 nations have agreed.

Researchers hope the guidelines will help to calm global discord over the sharing of financial benefits from natural products, such as potential drug targets — particularly those found in poor nations that are rich in biodiversity.

After marathon negotiations at the Seventh Meeting of the Conference of the Parties to the Convention on Biological Diversity (CBD) in Kuala Lumpur, Malaysia, representatives agreed on 20 February for guidelines to be prepared in time for approval at the next conference in Rio de Janeiro, Brazil, in 2006.

The guidelines are set to incorporate the idea of an internationally recognized certificate of origin, as a kind of passport for any scientific discovery. This will allow the source country and developer to follow the trail from its place of origin to its commercial exploitation.

Some researchers have expressed concerns about the extra regulations. "It is a bitter pill to swallow. But the moral argument is uncontested," says plant taxonomist Matthew Jebb, acting director of the National Botanic Gardens in Dublin, Ireland, who led the European Union's delegation in some of the talks. "We are reaping the legacy of more than 100 years of European domination."

The meeting rejected proposals from some poor countries for an extra treaty on access and benefit-sharing, as well as calls from rich nations for no extra rules.

Carlos Fernandez Ugalde, an environmental economist at Mexico's National Institute of Ecology and part of the Mexican delegation, says the outcome was satisfactory, but that it would be "a real challenge" to develop guidelines that would address the rights of indigenous people to traditional knowledge.

Delegates from Africa, Latin America and Asia complained about exploitation of genetic resources and failure to share the benefits. But drug companies say the CBD makes bioprospecting more trouble than it is worth in many countries.

The poorer countries also favour incorporating a discovery's place of origin into patent applications when the guidelines are implemented. But this is fiercely opposed by patent offices in rich countries such as the United States. ■

Society lifts publishing ban on nations facing US sanctions

Washington The American Chemical Society has announced that it is lifting a moratorium on the publication of scientific manuscripts from Iran, Libya, Sudan and Cuba.

The decision on 19 February came a week after a meeting with US government officials, who warned that accepting manuscripts from the four countries falls foul of current embargoes (see *Nature* **427**, 663; 2004). Robert Bovenschulte, who heads the society's publications division, says the group decided to lift the moratorium, which has been in place since November, after reviewing trade and constitutional law. "We believe it's the right thing to do," he says.

Bovenschulte says that several scientific publishers are working with the White House and Congress to reverse the government ruling, which they believe violates the US First Amendment invoking the right to free speech. He also says that the society is considering legal action, but only as a last resort.

"I think Iranian chemists will be happy to see that these restrictions have been removed," says Habib Bagheri, a chemist

from the Sharif University of Technology in Tehran, Iran, who is currently a visiting scholar at Purdue University in Indiana.

Gamers promise army the world for virtual training

Washington The US army has teamed up with a computer games company to create a simulation of the whole Earth.

In a bid to help soldiers train around the globe without travelling, army researchers are working with **There**, a company based in Menlo Park, California, that specializes in games set in realistic three-dimensional environments. Together they will build a



War games: software company **There** will create a three-dimensional world for the US army.

virtual model of the entire planet, using existing data about Earth's terrain. Robert Gehorsam, a senior vice-president at **There**, says that the product will model the real world as closely as possible.

The artificial world will help the army to practise intelligence work, patrols and planning, as well as encounters with civilians. A group of soldiers who served in Iraq will test the system in the spring; a final version is hoped to be in place by September. But the team has a long way to go — so far only part of Kuwait City has been modelled in detail.

Human pesticide tests 'given green light'

Washington A National Academy of Sciences panel has advised the US Environmental Protection Agency to accept data from pesticide companies that test their products on humans — but only when the tests have been conducted under strict ethical guidelines.

Some chemical manufacturers, claiming that dosage limits extrapolated from animal tests are too stringent, have submitted human test data when seeking approval for pesticides. But when the Environmental Protection Agency's science advisory committees couldn't decide on the issue, the agency asked the academy to look into

it and stopped accepting human test data. That suspension was overruled by a US appeals court in June.

The academy panel also advised that human tests should not be done simply to establish a lower acceptable dose under agency regulations without offering a public-health benefit.

But Richard Wiles, senior vice-president of the Washington-based Environmental Working Group, says: "The chemical industry will view the report as a green light to continue the highly unethical practice of dosing people with pesticides."

Drug trial in monkeys offers hope of SARS treatment

London A drug used to treat hepatitis C has been shown to relieve the symptoms of severe acute respiratory syndrome (SARS) and reduce infectiousness in monkeys.

In a study published this week in *Nature Medicine*, a team led by Albert Osterhaus of the Erasmus Medical Centre in Rotterdam, the Netherlands, tested the antiviral drug interferon- α on macaques.

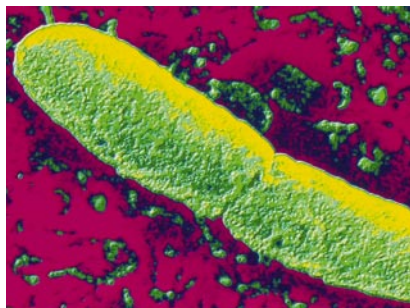
Monkeys were given doses of the drug and then infected with SARS. Two days later, their throats were almost free of viral particles, meaning that they exhaled fewer infectious particles — unlike untreated animals. "If the

same thing is seen in human patients, maybe the virus will spread less," says Osterhaus. The next step is to test the drug against SARS in humans, if another outbreak occurs.

Plague vaccine shows early promise

London British scientists are claiming early progress in developing a candidate vaccine against the bacterium that causes plague.

The scientists, led by Rick Titball, a microbiologist at the UK Ministry of Defence's laboratory in Porton Down, say their vaccine is based on proteins from the bacterium, *Yersinia pestis*, that activate the human immune system without causing



The plague bacterium *Yersinia pestis* could form the basis of a bioterror threat.

disease. Titball tested the vaccine on a small group of healthy volunteers and found it to be safe. His team now wants to conduct a large-scale efficacy trial involving several thousand volunteers.

There is currently no effective vaccine against plague, which could theoretically be used in a bioterrorist attack.

HIV enzyme boosts base amplification of artificial DNA

London Researchers led by Steven Benner of the University of Florida have designed an enzyme that allows non-natural DNA to replicate itself (A. M. Sismour *et al. Nucl. Acids Res.* 32, 728–735; 2004). The enzyme multiplies DNA that contains extra 'letters', or bases, beyond the naturally occurring cytosine, guanine, thymine and adenine. DNA with artificial bases has been around for several years but the strands could not reproduce themselves automatically.

The new enzyme, based on a polymerase produced by HIV, can string together DNA strands containing six nucleotide letters, rather than the usual four. DNA probes containing artificial bases are often used to detect viruses such as hepatitis C, and this discovery could radically speed up the production of these probes, Benner suggests.



Taking the voters' pulse

Political strategies and careers are built and broken on the results of opinion polls. But polls' apparently small margins of error can hide large uncertainties. Tony Reichhardt surveys the issues.

Before the Iowa caucuses kicked off the US presidential primary elections on 19 January, most polling organizations were predicting a virtual tie between the top four Democratic candidates. But when the results were in, two men — Senators John Kerry and John Edwards — were far ahead. Vermont governor Howard Dean, a hot favourite in some nationwide polls, was a distant third.

What looked like a serious polling blunder was more likely due to a last-minute surge of support for Kerry, and the quirky rules of the Iowa caucuses, which allow voters to switch their support after the voting begins. But it was another reminder that poll results often need to be taken with a large pinch of salt.

Public opinion polls have been growing in popularity ever since market researcher George Gallup pioneered them in the 1930s as a way to sell newspapers. Most polls claim a margin of error of only 3%. But people tend to miss the fine print describing everything — from question wording to refusals by subjects to be interviewed — that can skew the results. A typical Harris poll disclaimer concludes: “It is impossible to quantify the errors that may result from these factors.”

Jon Krosnick, a psychologist and political scientist at Stanford University, would go further. “That margin of error you hear about is an illusion,” he says. All it really guarantees is that the people sampled were statistically representative of the larger population. In fact, he says, there are many other sources of error, such as interviewers who mis-enter responses and respondents who

mis-hear questions or answer quickly just to get off the phone. “At some point, surveys need to stop publicizing that silly number,” Krosnick says.

Despite the theoretical problems, pollsters tend to do very well in predicting the outcomes of elections, particularly on the eve of election day when undecided voters make up their minds. By one count, 84% of the polls taken before the 2002 US Senate and gubernatorial elections differed from the actual vote by less than their theoretical margin of error.

Counted out

But pollsters live in perpetual fear of embarrassments such as the Iowa result. Or the 1992 British election, in which they wrongly predicted a Labour Party win. Or the 2002 French presidential contest, in which far-right candidate Jean-Marie Le Pen made a strong showing in the first round of voting, despite being counted out in pre-election polls.

To avoid nasty surprises, pollsters are always tweaking their methods. One hot area of research, says Krosnick, is aimed at determining which survey respondents are likely to vote. It's not as straightforward as it seems, as many who say that they plan to vote, do not. Pollsters know this, and sometimes use follow-up questions to get a better idea of what the respondent will do. Did you vote in the last election? Are you registered? Do you know where your polling station is? No one has worked out which filters work best, says Krosnick, but he and other researchers have found something curious: “The more people

you throw out of the sample — that is, the smaller the group that you pick as likely voters — the more accurate you get, even when you get too small to be representative of the country.”

Another option available to pollsters is to weight survey results to cancel out known or suspected biases in the sample, such as an under-representation of minorities or excess of Republicans. This is especially true of the tracking polls that have proliferated in recent elections. Produced by companies such as Zogby International, tracking polls now appear almost every day in newspapers and on the Internet. These snapshots of opinion generally use smaller sample sizes than weekly polls. Academic survey researchers say that they tend to be less accurate, despite the reader's perception that one poll is as good as another.

Weighting formulas, like those used to determine probable voters, are idiosyncratic and often proprietary. They are the black boxes of the polling business, with the exact formulas seldom being disclosed, says Michael Traugott, a professor of communications studies at the University of Michigan, Ann Arbor, and past president of the American Association for Public Opinion Research. He says they should receive greater public scrutiny, in part because tracking polls are often interpreted as real swings in opinion that can lead candidates to change their tactics or messages.

Although the public seems hungry for instant polls, it is increasingly reluctant to participate in surveys of any kind — a problem not just for pollsters, but for all survey



Polls apart: John Kerry (far left) surprised the pollsters by soundly defeating Howard Dean (left) in the Democratic primaries. Phone banks set up for mass marketing and political campaigns (above) have made potential poll respondents less willing to participate.

E. AMENDOLA/AP

researchers, from telemarketers to social scientists. Pollsters are having to work harder than ever to get the 1,000-person sample needed to achieve 95% confidence of being accurate to within 3%.

Some say that market researchers have poisoned the well for survey research. Answering machines and caller-identification technology have made it easier to avoid unwanted calls, and those who do answer the phone are less cooperative. By one estimate, only 35% of people reached by phone during the 2000 campaign answered pollsters' questions, compared with 65% in 1985.

Hanging up

This makes polling harder. But do falling response rates increase the risk of error? Probably not, according to one study, which showed that people who hang up on pollsters give similar answers to those who cooperate. Scott Keeter, of the Pew Research Center for the People and the Press in Washington DC, and his colleagues compared the results from a quick, five-day survey of adults who happened to answer the phone to those from a more rigorous, eight-week survey that tracked down and interviewed elusive subjects. Keeter found that attitudes across 91 categories varied

little between the easy-to-reach and hard-to-reach groups¹.

Even with 100% response rates and perfect knowledge of who will vote, pollsters would still find it hard to call the winner of an election as close as the 2000 US presidential race, where 48.6% of 105 million votes went to Al Gore, and 48.3% went to George Bush. To accurately predict a contest decided by only 0.3% would require surveying about 100,000 people. "And under current economic conditions, nobody's even talking about interviewing 10,000 people," Krosnick says.

Nobody, that is, except web-based pollsters. The Internet offers the benefit of potentially huge sample sizes at much lower cost than phone surveys. But most web surveys introduce a whole new bias. Instead of being picked randomly, respondents sign up to participate. As a result, "researchers have absolutely no idea where the respondents came from," Krosnick says. Even with weighting factors, it's just not possible to get a truly representative sample, he adds.

On the plus side, research done by Krosnick and others has shown that web respondents tend to be more careful and precise in their answers than phone respondents, who are more likely to misunderstand ques-

tions and give rushed answers. Web respondents also tend to be more honest, even on sensitive topics.

Knowledge Networks, a research firm in Menlo Park, California, founded in 1998 by two Stanford professors, combines the features of phone and web surveys. The company selects its sample through random digit dialling, which allows it to reach both listed and unlisted numbers, it then provides willing participants with computer access to answer questions on the web.

In last year's California gubernatorial election, Knowledge Networks flagged eventual winner Arnold Schwarzenegger's rise long before conventional pollsters did. The reasons for this are not clear, says Krosnick. "Maybe respondents were a little embarrassed to say that they were going to vote for a weightlifting actor."

Points of order

Once pollsters find willing opiners, there's the matter of what to ask and how to ask it. Surveys of all kinds are vulnerable to errors introduced by question order and wording². An analysis of polls during the 2000 campaign by Monika McDermott of the University of Connecticut, Storrs, and Kathleen Frankovic, director of surveys for CBS News in New York found that it even makes a difference whether the question "Who will you vote for?" comes at the beginning or end of the interview. The percentage of undecided voters dropped sharply when it came last, leading the researchers to conclude that placement of this question accounts for at least some of the variance in poll results³.

None of this would be so bad if the public knew about the inaccuracy and the bias. But Susan Herbst, a political scientist at Temple University in Philadelphia, doubts that they do. Under the guise of scientific objectivity, polls have diminished public involvement in the political process, she says. Expressions of public opinion that are harder to quantify, such as town meetings or letters to the editor, are now all but ignored. "Journalists have less incentive to highlight a political demonstration by 100 people when a professionally executed random sample survey on the same issue indicates that the demonstrators are a minority," she says.

And so the love-hate relationship with political polls continues. According to Trautgott, who has studied attitudes towards polling, the public is generally dismayed about the proliferation of polls. But the single most important factor in people's judgments of a poll's accuracy is whether it agrees with their own view. ■

Tony Reichhardt writes for *Nature* from Washington.

1. Keeter, S., Miller, C., Kohut, A., Groves, R. M. & Presser, S. *Public Opinion Q.* **64**, 125-148 (2000).
2. Schwarz, N. *Am. Psychol.* **54**, 93-105 (1999).
3. McDermott, M. L. & Frankovic, K. A. *Public Opinion Q.* **67**, 244-264 (2003).

Virtual plagues get real

Mathematical models incorporating ecological data are starting to be deployed on the front line in the battle against infectious disease.

Virginia Gewin talks to the number-crunchers who are spearheading the assault.

During Britain's epidemic of foot-and-mouth disease in 2001, the government culled some 4 million cattle, pigs and sheep. These drastic measures were introduced following the advice of a new breed of epidemiologist — those who deploy mathematical models of disease ecology to predict the progress of an outbreak and the probable effectiveness of different control strategies. In this case, the modellers crunched data on the distribution of farms, the size of herds, wind patterns and records of recent animal transport.

Controversy still rages over the decision to rely on culling alone, rather than using vaccination to try and reduce the slaughter¹. But the cull did succeed in bringing an end to an epidemic that threatened to send British agriculture into meltdown². Before the modellers recommended that livestock on infected farms be culled within 24 hours of foot-and-mouth being reported, and animals on neighbouring farms within 48 hours, the disease was spreading completely out of control.

Britain's experience with foot-and-mouth disease is a high-profile example of a growing trend to incorporate ecological models of disease in strategies to protect human and animal health. Ecological factors, experts are realizing, frequently underpin disease outbreaks.

In the future, climate change, encroachment into natural habitats and other consequences of human activity promise to change our exposure to infectious disease. And despite scepticism from some traditional epidemiologists, ecological models are showing up on the front lines of public-health systems



all over the world. "Mathematical modelling will be equally as important as anything that comes from the human genome in terms of its utility in public health," claims Andrew Dobson, an ecologist at Princeton University in New Jersey.

The field first began to attract attention in the late 1970s, thanks to the work of Roy Anderson of Imperial College in London and Robert May, then at Princeton and now president of the Royal Society in London. They created general models of disease spread based on the population dynamics of hosts and pathogens^{3,4}.

Since then, ecological modelling, in particular of the spatial dynamics of disease, has become increasingly robust. Recent models have revealed how outbreaks of both measles⁵ and dengue fever⁶ can spread in repeated waves from central foci of infection. In the latter case, researchers led by Donald Burke of the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, showed how waves of dengue infection radiate across Thailand from the capital, Bangkok, every three years or so, moving at a speed of almost 150 kilometres per month.

Dengue, which is a major problem in the tropics of Asia, Africa and North America, provides one of the clearest examples of the value of ecological modelling for public health. The dengue virus is spread by the mosquito *Aedes aegypti*, which also carries yellow fever. Many mosquito species breed in



marsh edges, or in transient puddles such as those in hoofprints, but *A. aegypti* prefers to lay its eggs in water-storage containers. With this knowledge in hand, Dana Focks, principal scientist at Infection Disease Analysis, a public-health research company in Gainesville, Florida, has developed an accurate method to estimate the size of adult mosquito populations by counting pupae in these containers.

Under control

Focks, who used to work with the US Department of Agriculture, has developed models that predict the risk of disease transmission from the availability of suitable mosquito breeding habitats and the degree of immunity to the virus in the regional human population⁷. These can be used to predict what types of water-storage bins should be covered or drained to keep

mosquito populations below the 'tipping point', beyond which a dengue epidemic becomes imminent.

Unfortunately, not all diseases have such straightforward, well-documented ecology. Malaria, for instance, is spread by more than 60 species of the *Anopheles* mosquito — for which there is little high-quality information on population dynamics.

In such cases, ecologists are trying to use secondary means of predicting mosquito distribution and abundance — such as satellite data and hydrological models of surface wetness. David Rogers, an ecologist at the University of Oxford, UK, is using satellite images to correlate seasonal climate parameters, such as annual rainfall and wet-season temperature, with predictions of the distribution of five of the six main mosquito vectors of malaria across Africa⁸.

It all adds up: mathematical models helped halt the UK foot-and-mouth epidemic by suggesting an appropriate culling strategy (left). Similarly, the spread of dengue fever (below left) can be pre-empted thanks to knowledge of the mosquito that causes its spread (below).



Longer-term climate cycles can also influence the prevalence of human disease. For example, the El Niño–Southern Oscillation (ENSO), a periodic disruption in trade winds, ocean currents and sea surface temperatures in the tropical Pacific, has been linked to outbreaks of cholera, dengue and hantavirus — which causes a deadly lung disease. Indeed, a team led by Gregory Glass of the Johns Hopkins Bloomberg School of Public Health has used satellite data to show how increased rainfall in the southwestern United States under ENSO conditions causes an increase in vegetation. This correlates with booms in populations of the deer mouse (*Peromyscus maniculatus*), which can pass hantavirus to people⁹. Terry Yates of the University of New Mexico in Albuquerque, a member of Glass's team, is now testing mathematical models designed to predict the risk of hantavirus transmission, which

are based on an ecological understanding of where infected rodents are likely to persist even in the absence of ENSO.

But it can be difficult to produce reliable predictions of disease incidence. Focks, for instance, is trying to use information on ENSO to develop dengue early-warning systems for urban Indonesia and Thailand. For Indonesia, unpublished results from his models produced reasonably accurate forecasts over a timescale of up to three months, but predictions for Thailand were not so successful.

Predictive power

Given the looming phenomenon of global warming, ecological modellers would dearly love to increase the reliability of predictions based on changes in climate. But at present, modelling the influence of climate change on the incidence of a disease such as malaria is difficult. Models based on the biology of the mosquito vectors suggest that, with global warming, malaria will spread into regions that are at present too cool to support a permanent mosquito population. Yet models that base their predictions purely on climate forecasts generally indicate that warming will have little impact on malaria prevalence⁸. "The state of the art of climate and infectious-disease modelling is improving, but hasn't yet reached a point where it could influence decision making," says Burke.

In other cases, modellers are frustrated by the sheer complexity of host and pathogen population dynamics. When asked by the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, to create a model to describe the spread of West Nile virus, Focks ultimately declined. "There were too many hosts, and not enough known about them, to make models like I did for dengue," he says. Indeed, West Nile virus, which in recent years has fanned out across most of the United States, has been found in 216 bird species, 30 other vertebrates and 49 species of mosquito.

These limitations may explain why some traditional epidemiologists remain unconvinced of the value of ecological modelling. The sceptics also question the extent to which ecological models can be used to devise viable disease-control strategies, given that decisions to fell forests, divert rivers or change the environment in other ways that may alter disease transmission are rarely made with public-health considerations in mind. "My personal opinion is that mathematical models will rarely be very useful," says David Morens, a medical epidemiologist at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland.

Sceptics are particularly concerned by claims from some ecologists, including newcomers to the field, that they can devise models to predict outbreaks of disease. Such claims have won funding from agencies including the US National Science Foundation and the National Institutes of Health.

The problem, according to some public-health experts, is that in few cases are the underlying ecological data good enough to give the resulting models any real predictive power. "It does a disservice to the field to mislead policy-makers that we can use models for predictive capability when we don't have the data," says Duane Gubler, now at the University of Hawaii in Honolulu, and formerly director of the CDC's Division of Vector-Borne Infectious Diseases.

Model behaviour

Gubler argues that ecologists now entering the field should look to emulate Fock's work on dengue. "His models are the best in the business, and the reason they are so good is because he understands the complexities of the ecology that goes into them, and he understands their limitations," Gubler says.

Despite the doubts, the World Health Organization is embracing a number of ecological models in its efforts to control water- and vector-borne diseases. Following the foot-and-mouth epidemic, Britain's Department of Environment, Food and Rural Affairs has incorporated mathematical modelling into its cadre of tools for protecting the health of livestock. And US institutions are beginning to follow suit. New Mexico's state government is using Yates's model to provide warnings of when environmental conditions may support a hantavirus outbreak.

Whether ecological modelling expands further to become an established public-health tool will depend upon the incorporation of better underlying data. But Burke, a self-proclaimed convert to the power of ecological models, believes that another major impediment to progress in the field is a lack of dialogue between mathematical modellers and traditional epidemiologists. "They don't talk to each other as much as you think they would," he says. If public-health officials are to meet their ultimate goal of predicting and preventing disease outbreaks, rather than simply responding to mitigate their toll, suggests Burke, this needs to change. ■

Virginia Gewin is a science writer in Corvallis, Oregon.

1. Keeling, M. J., Woolhouse, M. E. J., May, R. M., Davies, G. & Grenfell, B. T. *Nature* **421**, 136–142 (2003).
2. Woolhouse, M. et al. *Nature* **411**, 258–259 (2001).
3. Anderson, R. & May, R. M. *Nature* **280**, 361–367 (1979).
4. May, R. M. & Anderson, R. M. *Nature* **280**, 455–461 (1979).
5. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. *Nature* **414**, 716–723 (2001).
6. Cummings, D. T. et al. *Nature* **427**, 344–347 (2004).
7. Focks, D. A., Brenner, R. J., Hayes, J. & Daniels, E. *Am. J. Trop. Med. Hyg.* **62**, 11–18 (2000).
8. Rogers, D., Randolph, S. E., Snow, R. W. & Hay, S. I. *Nature* **415**, 710–715 (2002).
9. Glass, G. E. et al. *Proc. Natl Acad. Sci. USA* **99**, 16817–16822 (2002).

Intelligent plagiarists are the most dangerous

How should we tackle the increasing problem of researchers rewriting others' results?

Sir — In your News story “Plagiarism in Cambridge physics lab prompts calls for guidelines” (*Nature* **427**, 3; 2004) and your Editorial “Complacency about misconduct” (*Nature* **427**, 1; 2004), journals were criticized for not responding appropriately to plagiarism.

In my work as one of the editors of *Physica Scripta*, I have in recent years seen an increasing number of attempts at blatant plagiarism. For example, in 1973 I was one of five authors of a paper published in *Physica Scripta* (7, 241–249; 1973). Many years later an almost identical paper appeared in the *Indian Journal of*

Pure and Applied Physics (**36**, 273–279; 1998), where the main difference was that our names had been replaced by the names of two other authors.

However upsetting this may be, I believe that such naive plagiarism is not a significant threat to confidence in science, because such behaviour will ultimately have limited success.

What is worse, in my opinion, but was not discussed in these *Nature* articles, are cases where scientists rewrite previous findings in different words, purposely hiding the sources of their ideas, and then during subsequent years forcefully

claim that they have discovered new phenomena. Such ‘intelligent plagiarism’ is, unfortunately, often more successful because most scientists do not have either time or sufficient interest to carefully investigate where the original results came from.

As such misconduct seems to me to have recently increased within the scientific community, I think that a thorough discussion of these issues, in *Nature* or elsewhere, is urgently needed.

Lennart Stenflo

*Department of Physics, Umea University,
SE-90187 Umea, Sweden*

Overseas scientists still welcome in United States

Sir — I was pleased to see the headline of your Editorial “In praise of immigration” (*Nature* **427**, 181; 2004). However, it was distressing to read the standfirst, which says “officials of the [US] federal government don’t seem to recognize that the country’s scientific strength depends in large part on foreign talent”. Further, you go on to quote the National Science Foundation director, Rita Colwell, in support of this argument.

First, I must clarify that, although I am serving as chair of the National Science Board, the board has not yet had an opportunity to develop an official response to the *Nature* Editorial. However, I wanted to quickly provide my personal views to correct the potentially far-reaching and counterproductive impression conveyed by your Editorial.

Sadly, your Editorial seems to focus on a single, nine-word remark by Dr Colwell during a lengthy National Science Board press conference on the US scientific workforce. That remark, taken by itself, does not represent the view of the report presented by myself and other board members, including Dr Colwell. It is particularly frustrating that your article ignored a key theme of the Board’s report (see www.nsf.gov/nsb/documents/2003/nsb0369/nsb0369.pdf): the continued need for the United States to attract science and engineering students and professionals from other countries. This fact was stressed repeatedly during our presentation. Indeed, the National Science Board recommends that future US policies “while enhancing our homeland and national security, maintain the ability of the United States to attract internationally competitive researchers, faculty and students”.

Your editorial gives the false impression that US science-oriented officials are unconcerned about the impacts on international students and professionals of US efforts to enhance homeland security in the wake of 11 September 2001. Nothing could be farther from the truth. The board has publicly stated concerns with the frustrating — and in some cases damaging — impacts on the lives of individual scientists and engineers, and on the overall US science and technology enterprise, from the security-driven changes in our treatment of all international visitors.

The National Science Board has worked and will continue to work with US science officials, with Congress and with the White House, to reduce the burden on science and engineering students and workers arising from our national security needs. In October 2003, the National Science Board asked representatives from the White House Office of Management and Budget and the US Department of State to attend a board meeting for the express purpose of discussing our concerns about current and future US visa policy (www.nsf.gov/nsb/meetings/2003/nsb03138/nsb03138.htm).

We hope your readers recognize the commitment of federal science officials to ensuring that international researchers and students continue to feel welcome in the United States as partners in the US science and technology enterprise. At the same time, however, the federal government needs to take substantial steps towards increasing opportunities for US citizens to become more involved in science and engineering, in terms of both public awareness and potential careers that may directly or indirectly use these skills and knowledge.

Warren M. Washington

*National Center for Atmospheric Research, 1850
Table Mesa Drive, Boulder, Colorado 80305, USA*

HIV denialists will exploit any journal’s tolerance

Sir — Richard Smith in Correspondence (“Milton and Galileo would back *BMJ* on free speech” *Nature* **427**, 287; 2004) defends the rapid-response website section of the *British Medical Journal* (*BMJ*) — despite its use by those who deny that HIV is the cause of AIDS — on the grounds that free speech should always be allowed.

That’s not the real issue, however. The ‘HIV denialists’ exploit the reputations of any bona fide scientific journals that publish their opinions, in whatever form. Will a member of the public always be able to appreciate the subtle distinction between a peer-reviewed publication in a journal and an informal posting on the same journal’s website? One can readily imagine a layman (or even a head of state) being steered towards a *BMJ* rapid-response posting and thinking “It’s published in the *BMJ* — it must be accurate information”.

The denialists crave respectability for their maverick opinions, and anything that energizes them to continue their efforts to damage science and public health is to be deplored. Let them exercise their right to free speech on their own websites, not on one run by a respected medical journal.

John P. Moore

*Joan and Sanford I. Weill Medical College
of Cornell University,
Department of Microbiology and Immunology,
1300 York Avenue, W-805 New York,
New York 10021, USA*

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

The future of gene therapy

Balancing the risks and benefits of clinical trials.

**Marina Cavazzana-Calvo,
Adrian Thrasher and Fulvio Mavilio**

Gene therapy has the potential to treat devastating inherited diseases for which there is little hope of finding a conventional cure. In the late 1990s, our groups in Paris, London and Milan began treating children suffering from rare immune disorders (severe combined immunodeficiencies, or SCIDs). The successful treatment of the first patients was greeted with excitement when it was first reported in 2000 and 2002 (refs 1, 2). Sadly, this euphoria turned to alarm at the end of 2002, when two of the ten children treated in France developed leukaemia-like conditions³.

In all of these patients, the genetic defect was corrected by inserting a therapeutic gene into a 'disabled' retrovirus, known as a vector. This vector was then used to 'infect' bone-marrow stem cells taken from each patient before being injected back into their bloodstream, where it was hoped they would multiply into normal immune cells. As it turned out, the ability of these viruses to insert themselves into DNA was also responsible for activating a cancer-promoting gene. News of the leukaemia immediately raised concerns about using a therapeutic approach that may cause cancer at such an alarming frequency. Patient safety is, of course, the first and foremost concern of anybody trying to develop a new medical treatment. But it would be unfortunate if the future of gene therapy was decided by emotive issues rather than careful analysis of its risks and benefits.

The leukaemia cases generated enormous interest among scientists, regulators and the general public. Reactions from regulatory authorities in the United States and Europe varied widely (see Box, overleaf). Some asked clinicians to revise the eligibility criteria for future trials and to update procedures for obtaining informed consent from patients, whereas others imposed a general moratorium on trials involving the use of retroviruses. In the United Kingdom, clinical studies were never put on hold, whereas in Italy treatment for individuals was approved during 2003 only when there was an imminent threat to life. The combination of bad press, scepticism from colleagues, and mixed reactions from regulators has effectively thrown the field into recession.

The current 'gene-therapy-causes-cancer' mood and uncertainty about the effects of tighter regulations is discouraging scientists from starting new clinical trials, and scaring



Calculated risk: the use of viral vectors to deliver corrective genes to a patient can cause side effects.

investors and the biotechnology industry away from the field. In 2003 leading industrial players either closed their operations (Gene Therapy, Maryland) or redirected their efforts away from retroviral vectors (Cell Genesys, California). This is unfortunate, because in the absence of industrial investment it is unlikely that gene therapy will eventually deliver on its promises.

What can or should be done? As scientists, we are used to learning from crises and developing solutions to emerging problems. But restoring confidence in the future of gene therapy is going to be a tough sell — much of the debate is no longer about scientific concerns. We would like gene therapy to be seen, and treated, as any other experimental therapy, and that means recognizing the successes as well as the failures.

The successes

SCIDs are rare genetic failures in the development of the immune system that are fatal in the first years of life⁴. The patients treated in Italy suffered from adenosine deaminase-deficient (ADA⁻) SCID, which means that they lacked an essential enzyme involved in DNA metabolism. The patients treated in France and Britain suffered from X-SCID, caused by a defect in a gene on the X chromosome. For both diseases, transplantation of bone marrow from perfectly matched donors is the treatment of choice, although it is available to less than one

in three patients. For the others, transplantation from mismatched donors carries a 75% overall chance of survival, with a 15–20% risk of developing severe immunological complications and a 20–30% risk of early mortality⁵.

In the early 1990s, attempts to treat ADA-SCID with gene therapy achieved only partial success, owing to problems in transferring genes into the patients' stem cells. But recent developments in both vector and cell transplantation technology led to the successful treatment of both forms of SCID^{1,2}. Of the 18 SCID patients treated so far in Paris, Milan and London, 17 benefited from life-saving reconstitution of their immune functions for up to five years. All these patients are currently alive.

The failures

The optimism generated by what was considered to be the first true success of gene therapy turned into disappointment at the news of the two leukaemia cases. Genetic analysis of the malignant cells showed that in both cases the retroviral vector had inserted into, and activated, an oncogene called *LMO2* that is associated with childhood leukaemia. The activated oncogene was not the only cause of the malignancy, but was most likely the event that triggered it³.

The leukaemias came as a surprise, partly because none of the preclinical studies had shown any evidence of cancer in animals

treated with the same approach. Moreover, although scientists had always considered the possibility that gene insertion would activate oncogenes, no such event had been observed in more than a decade of clinical trials in humans involving large numbers of genetically modified blood cells⁶. The apparently high (15%) risk to X-SCID patients of developing malignant cells suggests that there are specific risk factors for this disease — one possibility being an association between the therapeutic gene and the activated oncogene^{3,7}. More research is needed into this question.

The trade-off

The patients who developed the leukaemias received a particularly high number of genetically modified stem cells. The issue of dose is an important one. By considering what we know about the likelihood of retroviruses inserting into active genes and the number of potential oncogenes in the human genome, a recent analysis estimates that up to one in every 10,000 modified cells might harbour a dangerous insertion⁸. Although these predictions need to be checked against actual clinical data, it is likely that the higher the number of cells given to a patient, the higher the probability of receiving one that is potentially malignant. This possibility poses a real ethical dilemma. The effectiveness of gene therapy has been limited for years by inefficient technology. The SCID trials changed that by increasing both gene transfer and cell transplantation efficiency. Scaling back this efficiency might reduce the risk of side effects, but will almost certainly lower the chances of success. Is there a sensible risk/benefit balance in such situations?

After the leukaemia cases occurred, scientists and regulatory authorities called for a halt to clinical experiments. In most countries, trials were allowed to resume after a temporary hold, on the basis that the potential benefits to patients outweighed the risks. Many argued that there is a need to develop new, safer vectors that avoid the problem of oncogene activation, and for more preclinical studies to enable better assessment of the risks. It is hard to disagree with these positions. Research must go on, particularly on vector design. Nevertheless, this work may take many years, and even the best animal model is far from being able to predict all of the possible risk factors when treating patients. This was certainly true in the X-SCID case.

In the meantime, we are left with a life-saving treatment that works for most patients, and several patients with no alternative treatments or with alternatives that carry even higher risks. Gene therapy, just like any other treatment, has side effects, and we have to deal with them — as we would do for any other treatment. Even with the risk of leukaemia, gene therapy is still a much better therapy than mismatched bone-marrow

Regulatory responses in Europe and the United States

United States

The FDA allows gene-therapy trials for X-SCID if no other therapy is available. Clinical hold on other stem-cell gene-therapy trials may be lifted after case-by-case review.

♦ www.fda.gov/ohrms/dockets/ac/03/minutes/3924M2.doc

United Kingdom

Approved clinical SCID trials are assessed on a case-by-case basis and are ongoing.

♦ www.doh.gov.uk/genetics/gtac/recommendationsGTAC-CSM.PDF

France

After a temporary hold, the French reopened clinical studies for X-SCID in January 2004.

♦ afssaps.sante.fr

Italy

Moratorium on any clinical trial involving the use of retroviruses until 31 December 2003. New ruling is currently awaited.

♦ www.iss.it/sitp/scf1/comu/index.html

Germany

After a temporary hold on all trials involving retroviruses, gene-therapy trials for SCIDs and other diseases restarted in February 2003.

♦ www.bundesaeztekammer.de/30/Ethik/80Themen/85KomSomGen

Europe

No Europe-wide regulations. Although experts argue that stem-cell gene-therapy trials should be allowed for life-threatening disorders after careful risk/benefit evaluation.

♦ www.emea.eu.int/index/indexh1.htm



Fresh hope: gene therapy can offer a realistic chance of survival to babies born with severe combined immunodeficiency.



Ready for action: custom-made gene vectors are held for use in a gene-therapy trial.

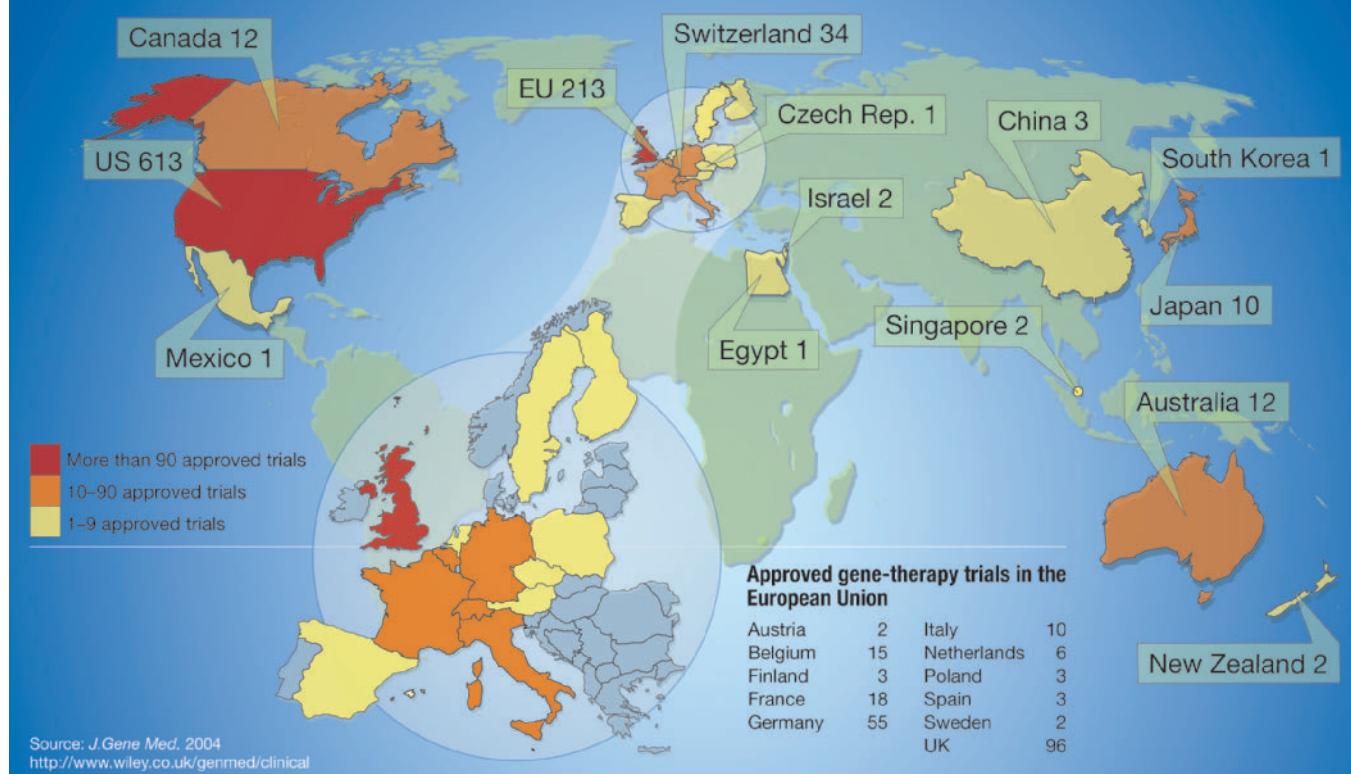
transplantation for SCID patients, and a fair assessment of the risk/benefit ratio should really be the only ethical criterion underlying the decision to use it. Ultimately, when assessing actual risk, there is no substitute for clinical trials on many patients. Delaying these trials would prevent researchers from assessing its full therapeutic potential, postpone its development into an effective therapy, and ultimately affect the right of patients to have access to a better treatment.

Prediction and monitoring of risks

Decisions on how to carry out further trials are complicated by the availability of technology to analyse genetically modified stem cells before and after they are given to patients. It has been proposed that using such analysis would prevent patients being given potentially malignant cells. Although regulatory authorities in Europe and the United States have not yet demanded this type of monitoring, the jury is still out on the issue. Could such 'molecular monitoring' provide useful information to clinical investigators and reduce the risks to patients?

As discussed above, there could be potentially dangerous gene insertions in all samples of genetically modified cells. Unfortunately, the molecular analysis destroys the cells. So screening all samples before transplantation, as argued by some⁹, is technically

Number of approved gene-therapy trials



impossible — each modified stem cell is unique, and taking 10% of the cells for analysis will tell you nothing about the other 90%. In addition, until we know more about what causes cells to become malignant, there is no evidence that detecting certain insertions would inevitably lead to leukaemia, for example. There are many factors that prevent cells with potentially dangerous insertions from developing into a malignant cell⁸, and the risk could vary greatly in different patients and for different diseases.

More realistically, molecular monitoring could be carried out during the follow-up phase of clinical trials. Such analysis could help to estimate the frequency of potentially harmful events, and provide a better risk assessment for future clinical trials. In the case of the French X-SCID patients, retrospective analysis of blood samples taken at regular intervals revealed how clones of the malignant cells grew and proliferated in the patients' blood³. This provided a wealth of information on the causes and development of the leukaemia. However, monitoring gene insertions has less value than we would like for predicting and diagnosing cancer in individual patients. For diagnostic purposes, proliferating cell clones cannot indicate cancer alone, which requires other clinical information. We therefore argue that mandatory molecular monitoring would put an unnecessary burden on clinical centres running gene-therapy trials, without

significantly reducing the risk of leukaemia. Instead, we recommend systematic archiving of bone marrow and blood samples from all patients undergoing stem-cell gene therapy, to allow retrospective analysis at any time.

Regulatory harmony

The varied responses from regulatory authorities add greatly to the uncertainty surrounding gene therapy. By creating a complex web of different rules in different countries, multicentre clinical trials become harder to plan and execute. Harmonization of legislation among European states, and between Europe and the United States, is urgently needed. Although talks are ongoing between the US Food and Drug Administration (FDA) and the European Medical Evaluation Agency (EMA) on this point, the picture remains bleak.

The EMA has no formal jurisdiction over early clinical trials, and individual European countries resist the idea of giving up national authority on this matter. The Gene Therapy Expert Group within the EMA has done an outstanding job in providing accurate information on the cancer-related risks, and sensible suggestions about the regulatory options¹⁰. However, unless measures are rapidly taken to harmonize the decision-making process of European states on gene-therapy regulation, it is unlikely that any agreement between the FDA and the EMA will have the beneficial

impact that investigators and industry — and ultimately patients — are waiting for.

Retroviruses are the only clinical tool currently available to introduce a permanent genetic modification into stem cells and to treat life-threatening conditions such as SCIDs. We believe it is essential to find a rational balance between feasibility, safety and efficacy when deciding on the clinical uses of these vectors, as well as when devising suitable regulations and guidelines. ■

Marina Cavazzana-Calvo is at the Immunology and Pediatric Haematology Unit, Hospital Necker, 75743 Paris, Cedex 15, France.

Adrian Thrasher is at the Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.

Fulvio Mavilio is at the Istituto Scientifico H. San Raffaele, Via Olgettina 58, 20132 Milano, and in the Department of Biomedical Sciences, University of Modena, Via Campi 287, 41100 Modena, Italy.

1. Cavazzana-Calvo, M. *et al. Science* **288**, 669–672 (2000).
2. Aiuti, A. *et al. Science* **296**, 2410–2413 (2002).
3. Hacein-Bey-Abina, S. *et al. Science* **302**, 415–419 (2003).
4. Fischer, A. *Lancet* **357**, 1863–1869 (2001).
5. Antoine, C. *et al. Lancet* **361**, 553–560 (2003).
6. Bonini, C. *et al. Nature Med.* **9**, 367–369 (2003).
7. Dave, U. P., Jenkins, N. A. & Copeland, N. G. *Science* **303**, 333 (2004).
8. Baum, C. *et al. Blood* **101**, 2099–2114 (2003).
9. Williams, D. A. & Baum, C. *Science* **302**, 400–401 (2003).
10. <http://www.emea.eu.int/index/indexh1.htm>

Acknowledgements

The authors report the shared opinion of the scientists and clinicians who developed and performed gene therapy of SCIDs in France, Italy and the United Kingdom, and in particular of Alessandro Aiuti, Claudio Bordignon, Alain Fischer, Salima Hacein-Bey-Abina and Maria Grazia Roncarolo.

From spears to speech

Could throwing spears have laid the foundations for language acquisition?

A Brief History of the Mind: From Apes to Intellect and Beyond

by William H. Calvin
Oxford University Press: 2003. \$26

Robin Dunbar

After a century of neglect, the mind has suddenly become an issue of evolutionary interest once again. Partly thanks to evolutionary psychology (by which I don't mean the pastiche that the media has largely been responsible for creating), the differences between human and non-human minds and their origins have begun to spawn a number of books. Notwithstanding the enthusiasm in the 1970s and 1980s for the similarities between humans and our primate cousins, both in popular culture and among academics, the fact is that humans are very different from even our ape sister species. William Calvin's latest book looks at how different we really are.

The essence of Calvin's argument is that the difference between humans and other animals comes down to what he calls "structured stuff" (that is, coordinated, structured task processing). One of the most obvious examples is the way we deconstruct sentences to expose their meaning. We can do this, he argues, because we evolved the capacity to coordinate fine-tuned movements in the context of throwing. The great revolution in human evolutionary history

stems from the shift from the older forms of heavy-duty hunting, mostly by dint of thrusting spears, to projectile hunting (throwing spears or using bows), which required careful aiming and much finer coordination.

Practice at these activities fine-tuned the neural machinery that allowed the delicate motor control required for speech and language. Much is made, in this respect, of the growing evidence for the brain's ability to coopt neural circuits. For example, the neural substrates for reading have different location in the brain in different individuals, as one might expect of a skill that does not have a long evolutionary history. This 'softwiring', as Calvin calls it, is clearly of major importance in human cognition.

I found the themes of the book, broadly speaking, congenial, and the account well informed and authoritative, as one might expect from a neuroscientist and science popularist of Calvin's stature. However, there are aspects of this particular book that I found less satisfying. Calvin's insistence on the importance of a gesturally based phase to language evolution does not, I think, make sense. Language is a parsing skill, and, even though parsing is a hierarchical process, it seems to me to be a very different kind of skill from that used in coordinated throwing. Manipulating concepts is not the same kind of activity as manipulating muscle

masses. Nor does the timing really work. The evidence, as Calvin himself notes, points to a period about 500,000 years ago as the likely timing for the origin of speech, if not full-blown language. But the archaeological record is very clear that real projectile-based hunting did not become widespread until the Upper Palaeolithic revolution, which kicked in around 50,000 years ago (perhaps a little earlier in Africa). The evolution of speech, then, pre-dates the fine muscle control of aimed throwing by a very wide margin.

I also found the writing style somewhat irksome. Of course, this is a book intended for a popular audience, and perhaps a certain kind of popular audience appreciates this kind of erratic, scattergun style of writing. I just find it irritating. The last chapter is an attempt to look into the future, our future. I found it largely unconvincing. There are substantive issues about the extent to which the natural limits to human cognitive flexibility might constrain our ability to cope with the stresses and strains of the modern urban jungle, or with the opportunities offered by the information highway. These issues are of genuine interest and real significance. But I saw none of them discussed here. ■

Robin Dunbar is at the School of Biological Sciences, Biosciences Building, Crown Street, University of Liverpool, Liverpool L69 7BZ, UK.



That's fighting talk: the neural control of the muscles required for firing arrows is similar to that needed to control speech.

D. LOPEZ-MILLS/AP

Trying to beat the clock

Rhythms of Life: The Biological Clocks that Control the Daily Lives of Every Living Thing

by Russell Foster & Leon Kreitzman

Profile: 2004. 320 pp. £20

Till Roenneberg

"We eat when we are hungry, sleep when we are tired." These are the first words of a new book that describes one of the most fascinating subjects of modern biology — the biological clock. In the triviality of these first words lies one of the reasons why science has neglected this subject for most of its history: it seems so banal that we get up when the Sun rises and go to bed when it sets.

Although Jean-Jacques de Mairan discovered in 1729 that daily rhythms in plants are independent of the Earth's light-dark cycle, the scientific birth of clock research lies in the middle of the nineteenth century, coinciding with the invention of the light bulb. One could argue that Thomas Edison is the godfather of circadian research, because it was with growing independence from sunlight that scientists became aware of the fact that the daily structure of our behaviour has a life of its own. Circadian research picked up pace throughout the twentieth century, culminating in the discovery of clock genes around the turn of the century.

In their book *Rhythms of Life*, Russell Foster and Leon Kreitzman succeed in walking a tightrope — although clearly understandable to the lay reader, the detailed description of the biological clock will also be fascinating for scientists. The authors take readers on a captivating journey, walking them through the history of clocks, both



The dark side of shift work: night workers suffer because they can't fool their circadian clock.

mechanical and biological, and guiding them through the spinney of terminology. They lead them on a hunt through tissues and cells, molecules and genes, in a search for the location of the biological clock, and for the parts that make it tick.

When shielded from the outside world, the biological clock runs free at its own pace. Depending on the organism and the individual, the period is slightly shorter or longer than 24 hours (hence circadian, meaning 'about one day'). The circadian clock is synchronized to the 24-hour daily

cycle by environmental signals or *zeitgebers* (a German term meaning 'time givers'), predominantly by light.

For most of us, light is the basis for viewing the world, yet there is more to light perception than just vision. This discovery is one of the most fascinating achievements of chronobiology, and Foster is the pioneer who made it possible. Vision relies on local contrast, and requires high spatial and temporal acuity. But the circadian clock, and probably many other biological functions, simply needs to detect whether there is light, and how much. In the course of deciphering the light input to the circadian clock, a completely new light-sensing machinery was discovered. The clocks of mice whose retinæ are completely devoid of rods or cones, which are needed for vision, still respond perfectly to light signals. It is a shame that the coverage of this fascinating story is kept relatively short — I hope that this will be the subject of another book.

After explaining the biological basis for daily timing, Foster and Kreitzman bring us back to civilization. With the help of light and the biological clock, our behaviours and all of the functions in our body are placed at the 'right' time of day. Humans are day-active, and as such have specialized senses (and a colourful appearance) that we share with other day-active animals — these characteristics are different from those of night-active creatures. But are we really safe from being active at the wrong time of day?

New in paperback

Freedom Evolves

by Daniel Dennett

Penguin, \$17

"As in his previous books, Daniel Dennett has rendered the difficult comprehensible and has ably defended a clear piece of intellectual territory." Melvin Konner, *Nature* **423**, 17–18 (2003).

Victorian Sensation: The Extraordinary Publication, Reception, and Secret Authorship of *Vestiges of the Natural History of Creation*

by James A. Secord

University of Chicago Press, \$22.50, £16

"A remarkably intricate picture of nineteenth-century British culture, in a book which is both coherent and agreeably written." David Oldroyd, *Nature* **409**, 285 (2001).

Hepatitis B: The Hunt for a Killer Virus

by Baruch S. Blumberg

Princeton University Press, \$17.95, £11.95

"The book is a gem.... It is essential reading for all aspiring scientists, the faceless bureaucrats who control the budgets for medical research, and devotees of the Research Assessment Exercise." Arie J. Zuckerman, *Nature* **417**, 900 (2002).

Love at Goon Park: Harry Harlow and the Science of Affection

by Deborah Blum

Berkeley, \$16

"A nuanced and brilliant evocation of a major figure in psychology who dared to challenge the orthodoxy of his time." Alison Jolly, *Nature* **420**, 741 (2002).

Science in culture

Mars the sublime

The first images from the Mars Express orbiter highlight the universal beauty of the wilderness.

Martin Kemp

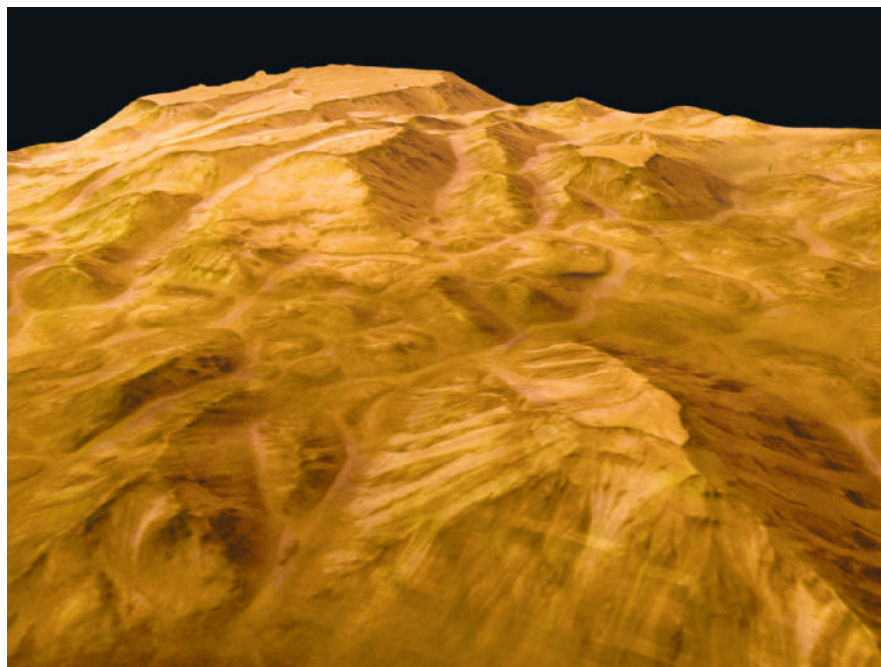
"It's just so beautiful as well as awe-inspiring... It's art and science and beauty and exploration all mixed up together". John Murray of Britain's Open University, one of the international team involved in the European Space Agency's current mission to Mars, is speaking about the first images to be produced by the High Resolution Stereo Camera (HRSC) on the agency's Mars Express orbiter. Looking at this view of the Grand Canyon of Mars (Valles Marineris), it is easy to see why.

But why should we find beauty in the barren topography of a planet that is so hostile to life? The answer lies in the interplay of our conditioned response to particular kinds of sublime landscape and the aims behind the image's generation. Our appreciation for painted landscapes that are awesomely terrible, rather than pastorally comforting, was the product of the Romantic imagination in the eighteenth century. Edmund Burke's *Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful* in 1757 set the tone.

Wilderness, particularly the vast tracts of newly explored territories in North America, became a source of aesthetic pleasure. James Fenimore Cooper's book *The Last of the Mohicans*, from 1826, plays out its tragic narrative against great vistas of mighty mountains, high plateaux and deep valleys. Our primitive and fundamental emotions seem to be more naturally expressed in savage topographies than in the artificial enclosures of cities.

The rhetoric of the sublime has long been integral to the imagery of the planets. Mars Express now presents us with images of the martian surface that are analogous to the aerial views on postcards that we might buy on a visit to the Grand Canyon. But the new pictures are not simply the high-tech equivalent of high-quality snapshots. We are witness to a superbly contrived visual model, the realism of which arises from the complex synthesis of the various data collected by the HRSC.

The data are the product of a 'linescan' camera that exploits the highest levels of optical and digital



This image of Valles Marineris from Mars Express shows that barren landscapes can be beautiful too.

sophistication. It works with nine independent image strips, each of which consists of 5,184 light-sensitive cells, orientated perpendicularly to the flight direction. Three of the channels register red, green and blue, and another is sensitive in the near-infrared range. Three strips record photometric measurements, and two are dedicated to stereo information, to bring about the complete three-dimensional modelling of the planet's surface in full relief. Additionally, the data are read out at variable frequencies to accommodate the ground velocity of the orbiter.

The HRSC's resolution of up to 10 metres per pixel from an altitude of about 260 km, the closest that Mars Express gets to the martian surface, is supplemented by a Super Resolution Channel, which produces more detailed pictures, of up to 2.3 metres per pixel. The former produces reference locations for the latter, just as we need maps of varied scales when we move from cross-country roads into the dense grids of cities. The

issue of resolution is crucial to the geological analyses that are the central goal of the explorations of Mars. Whatever detailed data might emerge from the Mars landers, fundamental judgements must be made on the basis of the gross topographies of the martian mountains and valleys seen in such images.

One attraction of the sublime landscapes that were painted around 1800 was that the grand topographies bore witness to the vast and ancient processes that shaped our planet. The geology of Charles Lyell and the visions of artists spoke the same language of huge transformation. When we look at the latest pictures of the dramatically scarred surface of the red planet, we bring to them the baggage of our aesthetic predispositions, to which scientists and non-scientists alike are powerfully subject.

Martin Kemp is professor of the history of art at the University of Oxford and co-director of Wallace Kemp/Artakt.

Modern man has created an indoor society, efficiently buffering environmental changes and, at the same time, choosing sustained dimness. Outside light intensities range between 10,000 and 100,000 lux (depending on cloud cover), whereas inside light rarely offers more than 100–400 lux. So the strength of the *zeitgeber* light, which is so important in synchronizing our biological clock, has decreased enormously as a result of industrialization. Humans have successfully conquered every available biotope, and the independence from sunlight now

allows us to conquer every temporal niche of day and night.

We experience the consequences, for example during shift work or jetlag. Over millions of years, our biological clock has learnt to interpret that we experience the brightest light during the day, independent of when we are active. But the biological clocks of shift workers (some 20% of the working population in the West) compare the dim light during the night shift with the bright light of day, and refuse to be fooled. As a result, shift workers and travellers

across time zones eat (and work) when they are tired and sleep when they are hungry.

We know that our industrialized lifestyle has many consequences. Living against our biological clock and shielding ourselves from daylight may be among the most important. It is time that the profound knowledge of clock research reaches everyone in our society. This book does an excellent job in this campaign.

Till Roenneberg is at the Institut für Medizinische Psychologie, Ludwig-Maximilians-Universität, Goethestrasse 31, Munich D-80336, Germany.

Cancer without disease

Do inhibitors of blood-vessel growth found naturally in our bodies defend most of us against progression of cancer to a lethal stage?

Judah Folkman and
Raghu Kalluri

Many of us may have tiny tumours without knowing it. In fact, autopsies of individuals who died of trauma often reveal microscopic colonies of cancer cells, also known as *in situ* tumours. It has been estimated that more than one-third of women aged 40 to 50, who did not have cancer-related disease in their life-time, were found at autopsy with *in situ* tumours in their breast. But breast cancer is diagnosed in only 1% of women in this age range. Similar observations are also reported for prostate cancer in men. Virtually all autopsied individuals aged 50 to 70 have *in situ* carcinomas in their thyroid gland, whereas only 0.1% of individuals in this age group are diagnosed with thyroid cancer during this period of their life. Therefore, it has long puzzled physicians and scientists why cancer develops and progresses to be lethal only in a very small percentage of people. The realization that a lot of us carry *in situ* tumours, but do not develop the disease, suggests that these microscopic tumours are mostly dormant and need additional signals to grow and become lethal tumours. So, what are these additional signals, and why are most of us protected from them?

The most likely answer is our body's inherent capacity to prevent the majority of these *in situ* tumours from recruiting their own new blood supply, thus preventing further growth owing to a lack of oxygen and nutrients. In the absence of a new supply of blood vessels by a process known as angiogenesis, an *in situ* tumour can remain dormant indefinitely. Paradoxically, it is proposed that angiogenesis itself is under the control of many genes in our body known to promote cancer (oncogenes) or suppress growth of tumours (tumour suppressors) — the same genes that are also involved in creating cancer cells, as encountered in the harmless *in situ* tumours. If the genes involved are the same, then why are most of us protected from the disease of cancer? To understand this better, we may have to define cancer as having two critical phases. In the first, acquisition of mutations, possibly due to genetic instability, leads to the transformation of normal cells in our body into cancer cells. This phase is not inherently lethal, and



Cancer caused by angiogenic tumours of the thyroid gland is a rare event, despite many of us carrying *in situ* tumours.

generally results in a microscopic tumour where the high rate of tumour cell division is balanced by cell death. The second phase involves a switch to the angiogenic phenotype, due to constant recruitment of new blood vessels, which converts the non-lethal *in situ* tumours into the expanding mass of tumour cells that is potentially lethal to an individual. Therefore, it is likely that there are critical governing factors that distinguish an individual's capacity to launch angiogenesis and enter a lethal phase of cancer.

This progression depends crucially on the balance between the *in situ* tumour's total angiogenic output and an individual's total angiogenic defence. The angiogenic output is contributed by growth factors such as basic FGF, VEGF, IL-8 and PDGF. An individual's defence is facilitated by endogenous angiogenesis inhibitors that are either associated with specific tissues or circulating in the blood. These inhibitors include thrombospondin, tumstatin, canstatin, endostatin, angiostatin, and interferon alpha/beta. Angiogenesis within the *in situ* tumour is probably initiated when the angiogenic stimulators overwhelm the host angiogenic defence. It is conceivable that such disruption in the angiogenesis balance is under the control of both the genetic make-up of any individual cancer cell and its microenvironment within the tumour. This would then explain why cancer in different individuals progresses at different rates, and also why some individuals enter the lethal phase of cancer and others do not, in spite of carrying cancer cells within their organs.

Such thinking is further fuelled by interesting occurrences in nature. For instance, there

is a very low incidence of solid tumours in patients with Down Syndrome, who circulate elevated levels of endostatin, an endogenous angiogenesis inhibitor, due to an extra copy of chromosome 21. An increased incidence of prostate cancer in patients with a specific polymorphism in endostatin is also noted. These examples suggest that either an increase or decrease in the angiogenic defence can alter the rate of cancer progression. Therefore, genetic control of physiological levels of endogenous angiogenesis inhibitors, such as thrombospondin-1, tumstatin and endostatin, may provide a last line of defence against the conversion of *in situ* tumours into a malignant

phenotype of cancer. Although drugs such as Avastin, which inhibits VEGF (the growth factor involved in blood vessel development), decrease the overall angiogenic output of the tumour and result in delay of disease progression, a viable additional strategy would be to boost the angiogenic defence of an individual by providing the natural angiogenesis inhibitors as part of anti-cancer therapy.

Molecular studies aimed at a better understanding of why many of us are protected from the disease of cancer may thus lead to identification of new, non-toxic drugs capable of converting cancer into a chronic manageable disease. If — as autopsies of trauma victims reveal — a majority of us have *in situ* tumours, then an important goal of future research will be to prevent disease in those individuals in whom the genetics favour progression of harmless *in situ* tumours into a lethal form of cancer. It is possible that one day, cancer might be treated as a chronic manageable disease, such as diabetes and heart disease. ■

Judah Folkman is at the Department of Surgery, Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. Raghu Kalluri is at the Center for Matrix Biology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA.

FURTHER READING:

- Black, W. C. & Welch, H. G. *N. Engl. J. Med.* **328**, 1237–1243 (1993).
- Hanahan, D. & Folkman, J. *Cell* **86**, 353–364 (1996).
- Hamano, Y. *et al. Cancer Cell* **3**, 589–601 (2003).
- Kalluri, R. *Nature Rev. Cancer* **3**, 422–433 (2003).
- Udagawa, T. *et al. FASEB J.* **16**, 1361–1370 (2002).

Turning the key on p53

David P. Lane and Peter M. Fischer

The p53 protein is among the most effective of the body's natural defences against cancer. News comes of a promising way of releasing the protein from its inhibitions to carry out its anti-tumour duties.

Almost exactly 25 years ago, a report appeared¹ describing the interaction between a cancer-causing virus and a host protein dubbed 53K. That entity, since called p53 and recognized as a tumour-suppressor protein because of its ability to prevent undue cell proliferation, has attracted ever-increasing attention: alterations in the *p53* gene are found in more than half of all human cancers, and the p53 protein plays a critical role in the host and tumour response to radiation and chemotherapy².

Research on p53 is now living up to its promise to deliver practical goods. In China, for example, it has delivered the first gene therapy approved for the treatment of human cancer³. Furthermore, from a paper published earlier this month, it seems that p53 can be selectively reactivated in tumour cells following systemic administration of a p53-activating peptide to tumour-bearing animals⁴.

Among the latest developments, however, those described in *Science* by Vassilev *et al.*⁵ are likely to prove the most important. The work of these authors suggests that small-molecule drugs can be designed to activate p53 by preventing the binding of a negative p53 regulator called Mdm2. Their report raises the prospects of new non-genotoxic general treatments for cancer, and that other protein interactions may finally become validated drug targets. The significance of the small-molecule approach is one of efficacy — such drugs are more likely to be absorbed and distributed in the body, and to enter cells and so reach their targets.

The p53 protein acts as a tumour suppressor by halting the cell cycle and triggering apoptosis (programmed cell death). It does so in cells subject to stresses such as DNA damage or low levels of oxygen, or to the oncogenic activation that promotes tumour development. Mice that lack p53 develop tumours at a young age, as do humans that inherit germline mutations in the gene. The same mice are remarkably resistant to radiation-induced death, establishing p53 as a rate-limiting component of the lethal response to DNA damage⁶. But how is p53 controlled so that it acts as a tumour suppressor but does not kill the organism? A huge number of proteins have been claimed to interact with and modulate p53, which is itself subject to a bewildering

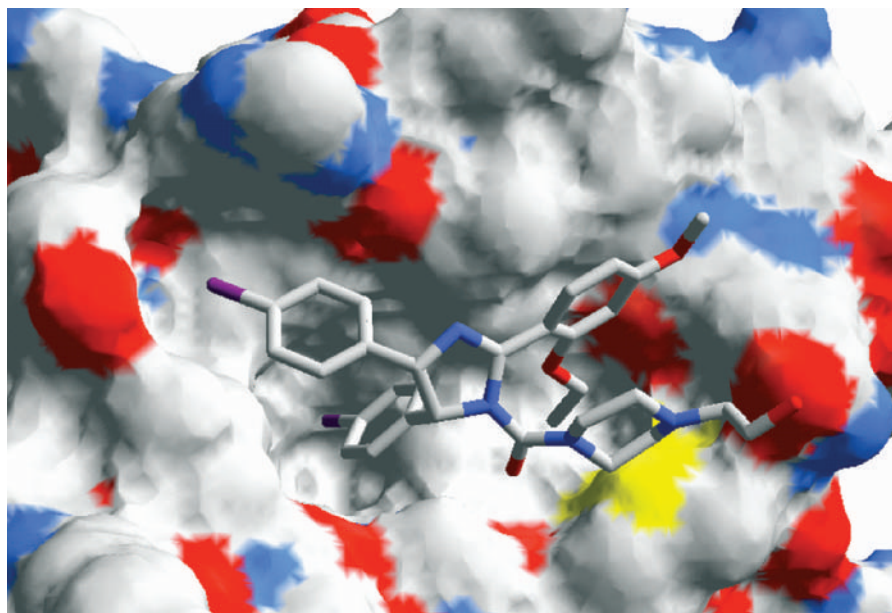


Figure 1 Inhibiting the inhibitor. A stick model of Nutlin-2, one of the small-molecule inhibitors designed by Vassilev *et al.*⁵, binds onto the same site on Mdm2 that p53 would occupy (the surface representation shown here is actually of the human equivalent of Mdm2). The three amino-acid side chains of p53 that interact intimately with Mdm2, and are mimicked by the Nutlins, are designated F19, W23 and L26 by the single-letter amino-acid code and position on the protein. (Image constructed from Protein Databank structure 1RV1.)

set of modifications. But one player, Mdm2, has been established as having a particularly central role in regulating p53 function⁷.

The Mdm2 protein binds to one end, the amino terminus, of p53 in a specific protein–protein interaction that can block p53 activity. Normally p53 works as a specific gene transcription factor, triggering the expression of a battery of genes that can initiate cell-cycle arrest or set off an apoptotic response. The Mdm2 protein inhibits this process by binding to p53 and blocking the binding of other transcription factors to p53's activation domain. In addition, Mdm2 can export p53 from the cell's nucleus and act as an enzyme that tags it for destruction in a cellular organelle called the proteasome. Genetic knockouts of the *Mdm2* gene in mice result in death at an early embryonic stage. But remarkably this outcome is completely dependent on p53: double-knockout mice, with both the *Mdm2* and *p53* genes disabled, are viable and fertile, although prone to cancer⁸.

These results led to the prediction that blocking the p53–Mdm2 interaction would be a way to activate p53 without having

to expose cells to oncogenes or genotoxic drugs. Such an approach could be valid in the half of human tumours that maintain expression of the normal *p53* gene. But enthusiasm for this strategy was tempered by three prejudices: first, that protein–protein interactions are difficult, maybe impossible, targets for small-molecule drugs; second, that activating p53 might be toxic to normal tissues; and third, that drug resistance would arise by mutation of the *p53* gene. From the results of Vassilev *et al.*⁵, it seems that all three of these objections might be overcome.

The first sign that the p53–Mdm2 interaction might be 'druggable' came from work, with small peptides, that established that Mdm2 binds with exquisite specificity to a 15-amino-acid segment of the p53 protein. Further studies established that highly potent peptide inhibitors of the interaction could be developed. When applied directly to cells or expressed as peptide fusion proteins, these Mdm2-binding peptides induced a p53 response in human tumour cells. The structure of an Mdm2–p53 peptide complex showed complete concordance with these biochemical and genetic studies, and

revealed that only three amino-acid side chains from p53 were buried in a deep pocket on the Mdm2 molecule⁹.

Since then, the race has been on to find the key for this lock. Vassilev *et al.*⁵ seem to have done so by identifying a family of synthetic compounds which they term Nutlins. They find that certain Nutlins can potentially displace p53 from Mdm2 and, as they show with newly solved X-ray structures, these compounds nestle into the p53-binding pocket on Mdm2 (Fig. 1). As predicted, the compounds effectively trigger p53 activation, cell-cycle arrest and apoptosis in tumour cells that retain normal p53, but are inactive in tumours that have mutant p53. When applied to normal cells they induce growth arrest but not apoptosis, and the growth arrest is reversible when the compounds are washed away. What is particularly exciting is Vassilev and colleagues' demonstration of the activity of Nutlins in animal xenograft models, in which they controlled the growth of tumours without apparent toxicity to the normal tissue. This specificity of p53 to kill tumour cells, but not normal cells, also seems to underlie the apparent safety of p53 gene therapy that has gained approval in China³.

What might be the basis of this phenomenon? A notable paper¹⁰ describes the subtle effect of a 'hypomorphic allele' of the *Mdm2* gene on the p53 response. Hypomorphism results in reduced but not completely curtailed production of a protein, and hypomorphic mice produce 30% of the normal level of Mdm2. Such mice are small, and show p53-dependent apoptosis of lymphoid cells. But they are nevertheless viable, implying that the effects of active p53 on cell viability vary from tissue to tissue and are by no means invariably lethal.

A further study identifies important differences between the p53 response in normal cells and in tumour cells¹¹. In normal cells, Mdm2 levels do not depend on p53's activity as a transcription factor, but in cancer cells this feedback pathway is active. Even more striking is the finding that in normal cells another tumour-suppressor protein, p14Arf, does not control Mdm2, whereas in tumour cells p14Arf is involved in negatively regulating Mdm2 and acts as a potent tumour suppressor. The active p53 pathway is therefore differently structured in normal cells and cancer cells. These remarkable observations help to explain the selectivity of the new molecules for inducing death in cancer cells, and at a deeper level they help to define the tumour state as one in which p53 signalling is controlled by p14Arf and by the Mdm2 feedback loop.

Two messages emerge from the paper by Vassilev *et al.*⁵ and the work on which it builds. The first is that protein-protein interactions can be successfully manipulated with small molecules. The second is that hypo-

morphic alleles may provide a much more informative biological model for drug action than simple all-or-nothing gene knockouts. The dream that basic research will deliver highly selective treatments for the common cancers seems closer than ever.

David P. Lane is at the Cancer Research UK Laboratories, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK, and at Cyclacel Ltd, which is developing small-molecule, anticancer drugs.

e-mail: d.p.lane@dundee.ac.uk

Peter M. Fischer is at Cyclacel Ltd, James Lindsay

Place, Dundee Technopole, Dundee DD1 5JJ, UK.

e-mail: pfischer@cyclacel.com

1. Lane, D. P. & Crawford, L. V. *Nature* **278**, 261–263 (1979).
2. Vousden, K. H. & Lu, X. *Nature Rev. Cancer* **2**, 594–604 (2002).
3. Surendran, A. *Nature Med.* **10**, 9 (2004).
4. Snyder, E. L. *et al. PLoS Biol.* doi: 10.1371/journal.pbio.0020036 (2004).
5. Vassilev, L. T. *et al. Science* **303**, 844–848 (2004).
6. Gudkov, A. V. & Komarova, E. A. *Nature Rev. Cancer* **3**, 117–129 (2003).
7. Iwakuma, T. & Lozano, G. *Mol. Cancer Res.* **1**, 993–1000 (2003).
8. Montes de Oca Luna, R., Wagner, D. S. & Lozano, G. *Nature* **378**, 203–206 (1995).
9. Chene, P. *Nature Rev. Cancer* **3**, 102–109 (2003).
10. Mendrysa, S. M. *et al. Mol. Cell. Biol.* **23**, 462–473 (2003).
11. O'Leary, K. A. *et al. Mol. Cell. Biol.* **24**, 186–191 (2004).

Astronomy

Out of the Dark Ages

S. George Djorgovski

Light from the most distant sources known, emitted when the Universe was only a billion years old, hints at a complex history of star and galaxy formation, and at their effect on the primordial gas around them.

The appearance of the first sources of light in the Universe ended the so-called cosmic Dark Ages, which had lasted a few hundred million years. The ultraviolet light from these sources also changed the physical state of the gas (hydrogen and helium) that fills the Universe, from a neutral state to a nearly fully ionized one. This was the era of reionization. Quasars — the brightest and most distant objects known (Fig. 1) — offer a window on the reionization era, because neutral hydrogen gas absorbs their ultraviolet light. On page 815 of this issue, Wyithe and Loeb¹ provide a new interpretation of the hydrogen-absorption spectra of the most distant quasars known. They conclude that a significant fraction of the cosmic hydrogen was still neutral about a billion years after the Big Bang. Coupled with other observations, their analysis adds to a growing body of evidence that the early history of galaxy formation was more complex than

was supposed even a couple of years ago.

When the Universe was about 380,000 years old, it underwent a phase transition, changing from an incandescent plasma containing the heat of the Big Bang, to a space filled with dark matter, energy and neutral gas. The glow of the primordial plasma is what we now observe as the cosmic microwave background (CMB). The Universe then entered the Dark Ages, with embryonic structures growing from the seeds of dark-matter fluctuations (fluctuations that are observed today as ripples in the CMB). After a few hundred million years, these condensations became dense enough for the first stars to form, leading to the appearance of the first galaxies and the growth of the massive black holes that are believed to power quasars.

As these first objects lit up, they also modified the gas between them, ionizing the hydrogen and making it transparent to

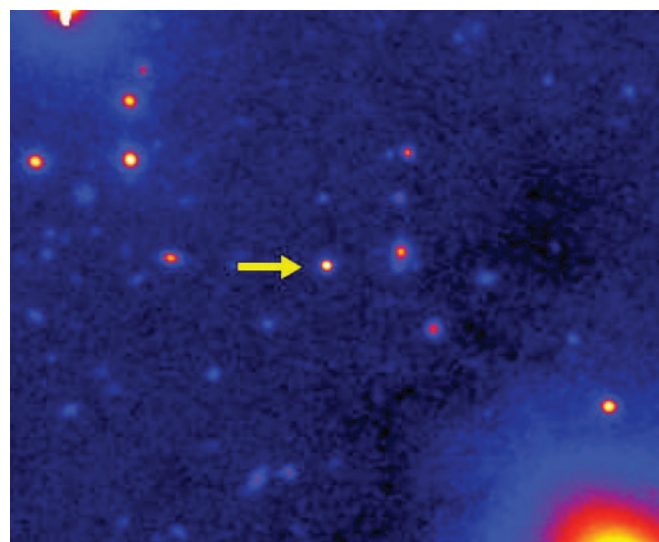


Figure 1 A long time ago in a galaxy far, far away. This image from the W. M. Keck 10-m telescope shows the quasar SDSS J1148 + 5251 as it was about 12.8 billion years ago: at a redshift of 6.41, it is currently the most distant quasar known. The spectra of ultraviolet light from this and other quasars at comparable redshifts have been used by Wyithe and Loeb¹ to constrain the reionization history of the Universe.

S. G. DJORGOSKI ET AL./CALTECH

ultraviolet light. Effectively, the Universe underwent another phase transition, from a neutral to an ionized state. Each of the primordial sources of light — most of them powered by young, massive stars, but some powered by the accretion of matter into growing black holes (the early quasars) — excavated a bubble of ionized gas, called a Strömgren sphere, in the otherwise neutral surrounding medium. When these bubbles began to overlap, reionization was complete. The reionization era is thus a cosmological milestone, marking the appearance of the first stars, galaxies and quasars.

The light reaching us now from distant quasars was emitted just before reionization was complete. Any neutral hydrogen remaining in the Universe at that point would have been a very effective absorber of this radiation at a particular set of wavelengths (extending up to 91.2 nanometres) known as the Lyman series. Even minuscule quantities of neutral hydrogen — as little as 10^{-4} or 10^{-5} by mass fraction in intergalactic clouds — can leave a tell-tale absorption signature in the spectra of quasars; this is the 'Lyman α forest' of absorption lines in quasar spectra. As the fraction of the neutral gas increases, the forest thickens; as the fraction approaches 100%, essentially all of the observed flux at wavelengths below the so-called Lyman α line (at a wavelength of 121.6 nanometres, in the restframe) is absorbed.

The existence of such an extended trough in the spectra of quasars was predicted in 1965 by Gunn and Peterson², and was indeed detected^{3,4} in the spectra of quasars at redshifts greater than 6. (Redshift is a measure of how much the Universe has expanded since the light was emitted; a redshift of 6 corresponds to a time roughly one billion years after the Big Bang in the favoured cosmological models.) Analysis⁵ of the spectra of the handful of quasars now known at such redshifts indicates that there is an abrupt change in the physical state of the intergalactic gas at this epoch. However, the actual fraction of the neutral gas is hard to constrain, because it involves measurements of fluxes that are very close to zero. Even with the best data available, all we can say is that the neutral fraction is at least 10^{-3} , which is not a very strong limit.

But a surprise was provided by the measurement of the polarized fluctuations of the CMB by the Wilkinson Microwave Anisotropy Probe (WMAP)⁶. These data indicated that the Universe was reionized at a redshift of 10 to 20 — that is, some 200 million to 500 million years after the Big Bang. How can this be reconciled with the quasar data, which suggest that the last stages of reionization occurred after one billion years?

Several authors^{7,8} have proposed that the Universe was in fact reionized twice. The first time was by a generation of massive luminous stars, and this gives rise to the WMAP

result. But the radiation from these stars and their eventual explosions as supernovae could have disrupted further star formation in their vicinity, leading to a hiatus during which ionized gas in the Universe might have recombined to become neutral again. A later generation of stars in young galaxies and quasars might have completed the reionization, at the end of the cosmic Dark Ages, when the Universe was about a billion years old.

Wyithe and Loeb¹ have put a clever new twist on the interpretation of the quasar spectra in which the Gunn–Peterson absorption effect is seen. They have used the extent of the ionized regions around the quasars themselves to derive a fresh constraint on the fraction of the neutral hydrogen at this time. A quasar will excavate its own bubble of ionized gas in the neutral gas around it. The extent of this bubble would depend on several factors, including the history of the quasar's own luminosity, and the neutral-gas fraction in the surrounding medium (more neutral gas would mean a smaller bubble radius). By making reasonable assumptions and modelling the possible growth history of

these quasars, Wyithe and Loeb conclude that, as the observed ionized bubbles around quasars seem small, the bubbles must have been embedded in a medium with a high fraction of neutral gas — as much as tens of per cent when the quasars turned on.

Coupled with the WMAP result, this implies that the Universe was indeed reionized at least twice. At any rate, the history of early galaxy formation and its effects on the primordial intergalactic medium seem complex. In our attempts to understand the reionization of the Universe, and the end of the cosmic Dark Ages, there are probably many more surprises to come.

S. George Djorgovski is at the California Institute of Technology, Pasadena, California 91125, USA.
e-mail: george@astro.caltech.edu

1. Wyithe, J. S. B. & Loeb, A. *Nature* **427**, 815–817 (2004).
2. Gunn, J. E. & Peterson, B. *Astrophys. J.* **142**, 1633–1642 (1965).
3. Becker, R. et al. *Astron. J.* **122**, 2850–2857 (2001).
4. Fan, X. et al. *Astron. J.* **125**, 1649–1659 (2003).
5. Fan, X. et al. *Astron. J.* **123**, 1247–1257 (2002).
6. Kogut, A. et al. *Astrophys. J. Suppl. Ser.* **148**, 161–173 (2003).
7. Cen, R. *Astrophys. J.* **591**, 12–37 (2003).
8. Wyithe, J. S. B. & Loeb, A. *Astrophys. J.* **586**, 693–708 (2003).

HIV

Replication trimmed back

Stephen P. Goff

A long-standing point of intrigue has been how certain non-human primates are resistant to HIV-1. The discovery in macaque monkeys of a protein that resides in mysterious cytoplasmic bodies holds the key.

Mammals have lived with retroviruses throughout their history. To avoid the detrimental effects of these RNA viruses, mammals have evolved a large number of genes to inhibit their replication. Certain non-human primates strongly 'restrict' HIV-1, blocking infection soon after this retrovirus enters the cell and before it starts making DNA from its RNA template during reverse transcription. Sodroski and colleagues (page 848 of this issue¹) have now isolated a gene, known as TRIM5 α , that is responsible for this restriction in rhesus macaques. The work uncovers key players in the early steps of virus replication, and may lead to new approaches towards inhibiting HIV-1.

The prototypical gene blocking early events in the retroviral life cycle, the *Fv1* locus, was first characterized in the 1970s. *Fv1* blocks particular mouse leukaemia viruses (MuLVs) soon after reverse transcription. Whether or not a virus falls victim to *Fv1* depends on its coat protein; in particular, a single residue at position 110 of the MuLV capsid coat determines this sensitivity. An important feature of the block is that it can be overcome if the virus is present in extremely high numbers. *Fv1* was found to

correspond to the capsid gene of a virus resident in the mouse genome, suggesting that it might compete with the incoming virus for critical host targets². How *Fv1* acts remains unclear, but there must be a specific interaction with the capsid protein of the incoming virus. Subsequently, many human cell lines have also been found to be resistant to particular MuLVs, an activity conferred by a gene named *Ref1* (refs 3, 4). Remarkably, the capsid is also targeted, and virus sensitivity is again controlled by residue 110. The *Ref1* gene of humans has not yet been isolated.

Excitement in the field increased with the realization that restriction was not limited to mouse viruses, but that many primates, including rhesus macaques and owl monkeys, restrict HIV-1 (refs 3, 5–9). The gene responsible, dubbed *Lv1* (ref. 7), in some species restricted HIV-1 but not the related monkey virus SIV_{MAC}. From a comparison of viruses containing portions of both HIV-1 and SIV genomes, the target of *Lv1* was pinpointed to be the HIV-1 capsid protein^{10,11}. Sodroski and colleagues¹ have now isolated the gene responsible from a rhesus monkey complementary DNA library by selecting for cells that manifest resistance to HIV-based vectors. The success of this approach is testi-

mony to the power of direct genetic selections. The active cDNA encodes TRIM5 α .

TRIM5 α , a protein resulting from differential processing of TRIM5 transcripts, passes all the tests as the main HIV-1 resistance factor present in primates¹. First, expression of the cDNA in human cells renders them resistant to HIV-1 infection, and the induced block occurs at the right stage of the viral life cycle. The rhesus TRIM5 α cDNA has potent antiviral activity, but its human counterpart does not, despite both proteins being expressed well. And whereas the rhesus gene product is active against HIV-1, it is only very weakly active against SIV_{MAC}, and not at all active against MuLVs. Furthermore, TRIM5 α blocks versions of the viruses that contain both macaque and human sequences — but only if the HIV-1 capsid region is present. Thus TRIM5 α targets the HIV-1 capsid protein, as expected. Finally, 'knock-down' of TRIM5 α , using small interfering RNA molecules that specifically target TRIM5 α transcripts for destruction, relieved the block against HIV-1 in rhesus cells, but had no effect on MuLV infection.

What is known about TRIM proteins? Not much. They are defined by a cluster of three different protein motifs: a RING motif, which is rich in cysteines and binds zinc; one or two so-called B boxes, which also bind zinc; and a coiled-coil domain that is probably involved in the formation of protein complexes. All individual TRIM proteins can probably aggregate with their peers. Some might also form alliances with other TRIM proteins.

A look through the human genome reveals at least 37 TRIM family members¹². The most famous member is PML, which can fuse with the retinoic acid receptor to cause some forms of leukaemia. PML also shows antiviral activity against many viruses. It resides in mysterious nuclear bodies of uncertain function, to which it may localize other proteins and selected messenger RNAs¹³. Each of the various TRIM proteins seems to localize to particular compartments within cells, forming discrete structures to which they entice other proteins. TRIM5 forms cytoplasmic speckles¹², consistent with the fact that it blocks HIV-1 infection before reverse transcription.

How might TRIM5 α block an incoming virus in the rhesus macaques? The work by Sodroski and colleagues¹ offers a few possible explanations. Infection of human cells by HIV-1 induces PML to move into the cytoplasm and bind the virus¹⁴; the ability of TRIM proteins to aggregate into large structures and to attract other proteins into them suggests that TRIM5 α might be binding and trapping the incoming virus in the rhesus macaques. Another possibility is that TRIM5 α could influence whether, and how, the capsid is modified, and thus its localiza-

tion. One such modification, reported for PML, is the addition of a 'SUMO' group. There are hints that the capsid proteins of MuLV and HIV-1 are indeed SUMOylated, and that this may be required for infection. Alternatively, TRIM5 α might interfere with the normal uncoating of the viral RNA core that precedes reverse transcription. This could involve interactions with cyclophilin A, a protein thought to affect the stability of the HIV-1 core. Indeed, there are strong hints that HIV-1 restriction is modulated by cyclophilin A (refs 10, 15).

A final possibility is that TRIM5 α initiates the degradation of the incoming HIV-1 particle. Other TRIM5 isoforms, including TRIM5 δ , behave as ubiquitin ligases¹⁶ — they add ubiquitin to other proteins, which labels them for destruction by the proteasome, the cell's disposal system. Perhaps TRIM5 α could transfer ubiquitin directly to the capsid. In support of this idea, treating human cells with proteasome inhibitors increased their susceptibility to HIV-1 infection¹⁷. Time will tell which of these ideas, if any, pan out.

Several straightforward questions arise. Does TRIM5 α bind directly to the HIV-1 capsid protein? Does it cause its relocation, modification or degradation? Does TRIM5 α comprise a subunit of a ubiquitin ligase — or of a similarly acting SUMO transferase? Is reverse transcription itself blocked by TRIM5 α , or is the resultant DNA

degraded after synthesis? To what else might TRIM5 α bind? Are TRIM proteins somehow involved in the mechanism of action of the *Fv1*, or the human *Ref1*, resistance genes?

All we can say for certain is that right now many laboratories are working furiously to be the first to find the answers to these questions. When the mechanism of action of TRIM5 α is uncovered, the next goal will be to recreate its effects in a therapeutic treatment. ■

Stephen P. Goff is at the Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York 10032, USA. e-mail: goff@cancercenter.columbia.edu

1. Stremlau, M. *et al. Nature* **427**, 848–853 (2004).
2. Best, S., Le Tissier, P., Towers, G. & Stoye, J. *Nature* **382**, 826–829 (1996).
3. Towers, G. *et al. Proc. Natl Acad. Sci. USA* **97**, 12295–12299 (2000).
4. Besnier, C. *et al. J. Virol.* **77**, 13403–13406 (2003).
5. Hofmann, W. *et al. J. Virol.* **73**, 10020–10028 (1999).
6. Munk, C., Brandt, S. M., Lucero, G. & Landau, N. R. *Proc. Natl Acad. Sci. USA* **99**, 13843–13848 (2002).
7. Cowan, S. *et al. Proc. Natl Acad. Sci. USA* **99**, 11914–11919 (2002).
8. Besnier, C., Takeuchi, Y. & Towers, G. *Proc. Natl Acad. Sci. USA* **99**, 11920–11925 (2002).
9. Hatzioannou, T., Cowan, S., Goff, S. P., Bieniasz, P. D. & Towers, G. J. *EMBO J.* **22**, 385–394 (2003).
10. Dorfman, T. & Gottlinger, H. G. *J. Virol.* **70**, 5751–5757 (1996).
11. Owens, C. M. *et al. J. Virol.* **77**, 726–731 (2003).
12. Reymond, A. *et al. EMBO J.* **20**, 2140–2151 (2001).
13. Borden, K. L. *Mol. Cell. Biol.* **22**, 5259–5269 (2002).
14. Turelli, P. *et al. Mol. Cell* **7**, 1245–1254 (2001).
15. Towers, G. J. *et al. Nature Med.* **9**, 1138–1143 (2003).
16. Xu, L. *et al. Exp. Cell Res.* **288**, 84–93 (2003).
17. Schwartz, O. *et al. J. Virol.* **72**, 3845–3850 (1998).

Earth science

Through thick and thin

Neil M. Ribe

The sea floor around the Hawaiian island chain is unusually shallow. New seismic evidence suggests that this up-raised 'swell' is partly due to heating and thinning of the lithosphere beneath.

Earth scientists revere the Hawaiian islands as the inspiration for some of the most fundamental concepts of their discipline. More than 40 years ago, J. T. Wilson suggested that the linear increase in the age of the islands along the chain reflected the steady motion of the Pacific plate over a source of heat, or hotspot, fixed in the mantle beneath¹. W. Jason Morgan subsequently proposed that hotspots were in fact the tops of mantle plumes — rising currents of hot buoyant material originating deep in the mantle². Meanwhile, ship-board surveys of the Pacific ocean floor had mapped out the Hawaiian 'swell', a large area of anomalously shallow sea floor that surrounds the Hawaiian chain (Fig. 1). What created the Hawaiian swell? The fact that it doesn't simply collapse under its own weight indicates that there must be an upward force that counteracts the pull of

gravity. On page 827 of this issue, Xueqing Li *et al.*³ weigh in with new data and fresh insight on the origin of the swell-supporting force.

There are two principal hypotheses. The first is lithospheric thinning⁴, or 'rejuvenation' (Fig. 2a). Here attention focuses on the lithosphere, the cold and mechanically stiff outermost layer of the Earth that includes the crust and part of the upper mantle. Under oceans, lithosphere a few kilometres thick is continuously created at mid-ocean ridges, and gradually thickens as it moves away and cools in contact with the ocean. The lithosphere under the Big Island of Hawaii, for example, has been cooling for some 90 million years, making it about 100 km thick. According to the rejuvenation model, however, normal oceanic lithosphere that happens to move over a hotspot is thinned, or rejuvenated, during the time it remains

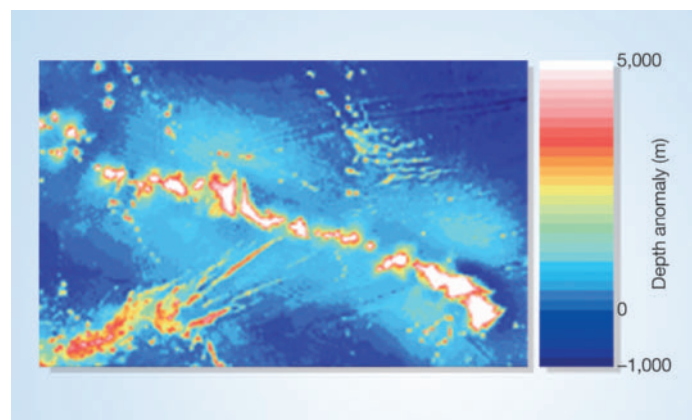


Figure 1 The Hawaiian swell. The sea floor surrounding the chain of islands is anomalously shallow, relative to normal sea floor of the same age, over an area about 1,200 km wide and 3,000 km long. The maximum height of the elevated region (excluding the islands themselves) is about 1,400 m.

over the hotspot — about five million years in the case of Hawaii, based on the distance over which the initial uplift of the swell occurs. The lithosphere then slowly thickens again by conductive cooling as it moves away from the hotspot. The result is an elongated trough at the base of the lithosphere, filled with rock that is hotter and less dense than the colder normal lithosphere to either side. By Archimedes' principle, this lighter material pushes the thinned lithosphere upwards to create a swell.

The rejuvenation model explains qualitatively the rapid uplift and the subsequent slow decay of the Hawaiian swell moving along the island chain to the northwest. It is also consistent with the observed anomalies in the gravitational potential over the swell, which imply that the compensating upward force originates at shallow depths. But the model does not propose a credible mechanism by which a hotspot can thin the fast-moving lithosphere by several tens of kilometres in a few million years.

In contrast to the purely static rejuvenation hypothesis, the second, 'dynamic support' model proposes that the swell is held up by the flow associated with an ascending

mantle plume. On reaching the base of the lithosphere, the buoyant plume material spreads laterally, forming a broad 'pancake' which pushes the lithosphere upwards (Fig. 2b). A simple physical model for the gravitational spreading of viscous fluid poured onto a moving plate^{5,6} predicts that the pancake will have a broad quasi-parabolic shape that agrees well with the shape of the Hawaiian swell. The model also predicts quantitatively the swell's rapid uplift and subsequent slow decay. But because the compensating density anomalies are below the base of normal lithosphere, the model has trouble accounting for the observed gravitational-potential anomalies.

Which of these models is correct? Li *et al.*³ propose that the answer might be both. They have used variations in the travel times of seismic waves generated by earthquakes to create a topographic map of the base of the lithosphere beneath the Hawaiian swell. The waves used are of two types: compressional (P) waves, with particle motion along the direction of propagation (like sound waves in air); and shear (S) waves with transverse particle motion. Li *et al.* start from the well-known principle that, when a seismic wave

of a given type (P or S) arrives obliquely at a boundary, some of its energy will be converted into a wave of the other type. The time required for this new wave (the converted phase) to reach a seismometer at the Earth's surface is a measure of the depth of the boundary at the 'piercing point' where the conversion took place. Using S-to-P converted waves, Li *et al.* have shown that the thickness of the lithosphere decreases from 100–110 km beneath Big Island to 50–60 km beneath the island of Kauai, 500 km to the northwest. The thinned region is about 300 km wide.

These results³, together with basic fluid-mechanical principles, suggest a new hybrid model for the origin of the Hawaiian swell — 'dynamic thinning' (Fig. 2c). In this picture, an ascending mantle plume strikes the base of normal oceanic lithosphere and spreads out laterally, as happens in the dynamic support model. However, as the plume material is carried northwest by the moving lithosphere, its hottest central portions begin to undergo thermal convection, induced by conductive heat loss to the cooler lithosphere above. Such convection — which may take the form of 'rolls' aligned with the plate motion direction^{7,8} (Fig. 2c) — transfers heat vertically much more efficiently than does conduction alone⁹. It will therefore erode and thin the lithosphere, producing a trough filled with hot plume material whose amplitude increases the further it gets from the hotspot. Accordingly, the swell as a whole will be supported dynamically by material in motion. But the material compensating the swell's inner portion will be substantially shallower than predicted by the traditional dynamic support model.

Following on from Li and colleagues' work, perhaps the most important task is to extend their analysis to greater distances

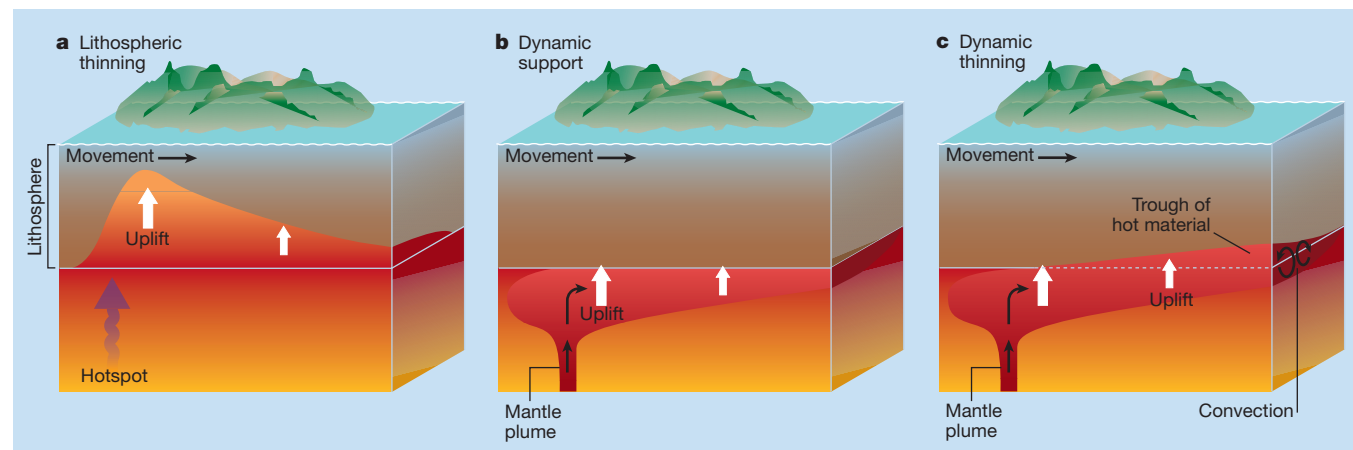


Figure 2 The origin of the Hawaiian swell. Two reasonably successful but incomplete models have been proposed. a, The first is the 'rejuvenation' model, in which the swell is compensated by buoyant material in a trough at the base of the lithosphere. The trough is created when heat from a fixed hotspot thins the moving lithosphere. b, According to the 'dynamic support' model, the swell is supported by hot material from an ascending

mantle plume, which spreads below the base of the lithosphere in the form of a broad pancake. c, Li *et al.*'s³ results support a hybrid of the two: the 'dynamic thinning' model. The swell as a whole is compensated dynamically, but convection currents within the hottest central part of the pancake cause the lithosphere to thin over an area about 300 km wide, starting about 100 km downstream from the vertical plume conduit.

along the Hawaiian island chain, with the help of new seismometers deployed further to the northwest. This might reveal the rethickening of the lithosphere that should eventually occur as the plume material cools. A second task is to test the conclusions of Li *et al.* against new measurements of heat flux on the Hawaiian swell. Thermally induced thinning of the lithosphere to 50–60 km should produce measurable anomalies of the heat flux at the ocean floor, starting about 10 million years later, or after about 900 km of motion by the Pacific plate. However, attempts to detect such anomalies¹⁰ have so far been unsuccessful. It seems the Hawaiian swell still has much to teach the Earth scientist. ■

Neil M. Ribe is at the Institut de Physique du Globe de Paris, 4 Place Jussieu, Paris 75252 Cedex 05, France.

e-mail: ribe@ipgp.jussieu.fr

1. Wilson, J. T. *Can. J. Phys.* **41**, 863–870 (1963).
2. Morgan, W. J. *Nature* **230**, 42–43 (1971).
3. Li, X., Kind, R., Yuan, X., Wölbern, I. & Hanka, W. *Nature* **427**, 827–829 (2004).
4. Detrick, R. S. & Crough, S. T. *J. Geophys. Res.* **83**, 1236–1244 (1978).
5. Olson, P. in *Magma Transport and Storage* (ed. Ryan, M. P.) 33–51 (Wiley, New York, 1990).
6. Ribe, N. M. & Christensen, U. R. *J. Geophys. Res.* **99**, 669–682 (1994).
7. Moore, W. B., Schubert, G. & Tackley, P. J. *Science* **279**, 1008–1011 (1998).
8. Ribe, N. M. & Christensen, U. R. *Earth Planet. Sci. Lett.* **171**, 517–531 (1999).
9. Sleep, N. H. *J. Geophys. Res.* **99**, 9327–9343 (1994).
10. Von Herzen, R. P. *et al. J. Geophys. Res.* **94**, 13783–13799 (1989).

Ion transport

Spot the difference

David C. Gadsby

Common wisdom holds that ion channels and ion pumps, with their obvious functional differences, should have visibly dissimilar structures. It seems that's not necessarily so.

The two kinds of membrane protein that control movements of ions such as Na⁺, K⁺ and Cl[−] into and out of cells have very different jobs. Ion channels let these charged atoms flow down chemical and electrical gradients; pumps build up the gradients. Hence the view that channels and pumps must be cut from different cloth.

In a startling dismissal of that idea, on page 803 of this issue Accardi and Miller¹ show that a protein called CLC-ec1, a prokaryotic forebear of the large CLC family of eukaryotic Cl[−] channels, isn't a channel as presumed^{2–5}. After reconstituting purified protein in an environment mimicking a cell membrane, the authors show that CLC-ec1 is an exchange pump that can harness an energetically favourable gradient of protons to move Cl[−] ions against their electrochemical gradient. This reclassification would excite only diehard transport specialists were it not for the fact that a high-resolution X-ray structure⁵ of CLC-ec1 has proven an illuminating model for understanding how CLC Cl[−] channels work. So we're forced to conclude that we can't yet distinguish a channel from a pump by simply staring at a crystal structure.

But what are the functional characteristics of channels and pumps that might demand structural differences? Ion channels need a membrane-spanning pore through which ions diffuse, and a gate that opens only as needed (lest perpetually open channels dissipate the gradients upon which their function depends). The pore of a selective ion channel must include a selectivity filter that allows speedy passage — 10 million to

100 million ions per second through the open pore — of the appropriate ions while keeping others out⁶. Ions move more slowly through pumps than through channels. For instance, the ATP hydrolysis-driven sodium–potassium pump, a so-called 'primary' transporter, exchanges three intracellular Na⁺ ions for two extracellular K⁺ ions only 100 times per second. The secondary transporter CLC-ec1 exchanges two Cl[−] ions for a single proton¹, up to 100,000 times per second^{1,7}.

In principle, the difference between an ion channel and an ion pump, or transporter, boils down to this: an ion channel needs no more than a single gate, and a pump needs at least two gates that ought never to be open at the same time (Fig. 1, overleaf). A practical limit on 'never' is set by the ratio of throughput rates for uphill versus downhill ion transport. A pump such as the sodium–potassium pump moves roughly 100 ions per second uphill, but would let 10 million ions per second flow downhill if both gates were open, so the probability of finding both gates open must be far less than one in 100,000 for the pump to be useful. To achieve such low probabilities, the pumps ensure that one gate closes before the other opens, temporarily trapping ions inside the protein (Fig. 1b). Although secondary transporters such as CLC-ec1 tend to have somewhat faster transport rates than the sodium–potassium pump, the same principle applies.

A CLC-ec1 exchange pump might therefore differ from a CLC Cl[−] channel by little more than an additional gate. So what constitutes a gate? Although, in theory, a



100 YEARS AGO

Studies in Heterogenesis. By H. Charlton Bastian. Heterogenesis means, in these studies, the *per saltum* origin of forms of life from other quite different forms, e.g. of a ciliated infusorian from a rotifer's egg, or of a sun-animalcule from a chlorophyll corpuscle. It is long since Dr H. Charlton Bastian first suggested this heresy; and many years of industrious observation have resulted in this large and expensive volume describing and (with 815 figures) illustrating those cases in which the author thinks he has detected the heterogenetic process at work. One cannot but admire the doggedness with which Dr Bastian has persisted — *contra mundum* — in maintaining his thesis; and even those who feel quite sure that he has misinterpreted what he saw may find it interesting to discover by repetition of his experiments what did actually occur and was actually photographed. Others, again, who would not turn round to look at slides supposed to demonstrate that the egg of a rotifer may resolve itself into infusorians or into one large ciliate, may be more tolerant of the suggestion that Protistan evolution is still going on, retracing some of its ancient steps, or making new ones. It may be that Proteus still frisks a little among the Protists, or that there are mutations among unicellulars just as among De Vries's evening primroses.

From *Nature* 25 February 1904.

50 YEARS AGO

I wish to direct attention to the large amount of scientific work carried out at the public expense but only reported in brief summary terms or published after long delay... It seems that an example is furnished by some of the work carried out under the Colonial Research Council. Under that Council is a committee dealing with insecticides which met first in January 1947. [This work] is lavishly financed and has received 7.5 per cent of the £12,000,000 allocated to the Colonial Research Council up to 1953, namely, £900,000. As most of the work carried out by the Insecticide Committee appears not to be very expensive, we conclude that the air spraying has cost about three-quarters of a million pounds. It is difficult to see why work should be done and paid for, if it is not made available or if it is to be published after it is out of date.

From *Nature* 27 February 1954.

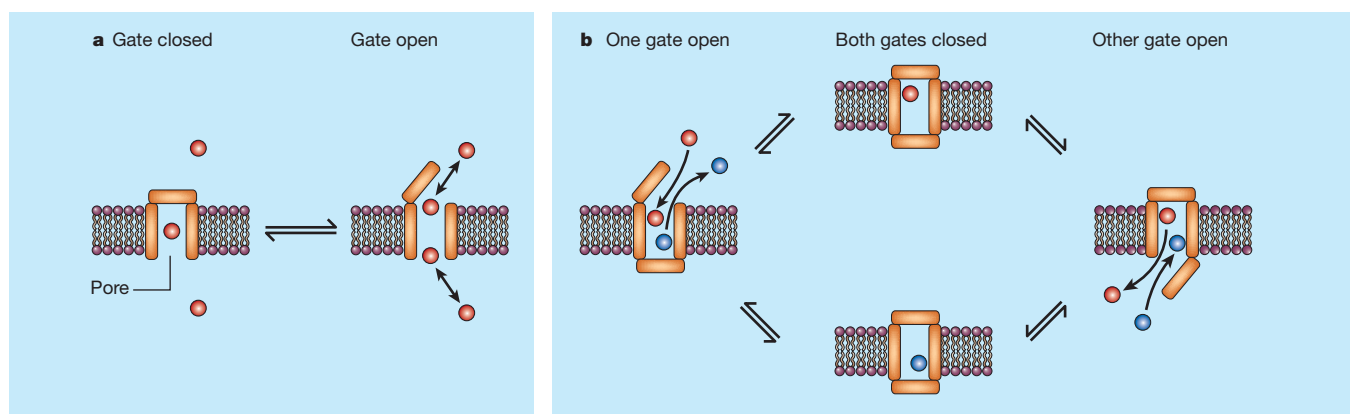


Figure 1 Functional differences between ion channels and ion exchange pumps. **a**, Ion channels need only a single gate, which can be open or closed. When closed, ions (red circles) cannot pass through the pore but when open, ions can enter and flow through freely. **b**, By contrast, ion-exchange pumps need two gates, which are never both open. Each gate can open only if the other is closed, to prevent unrestricted ion flow. When one

gate is open, one type (red) of ion enters, and another (blue) leaves. Continuing clockwise, closing the gate temporarily traps red ions inside the pore, until the second gate opens to release them to the other side of the membrane and allow blue ions in. Clockwise exchange of red for blue ions across the membrane is completed after the second gate closes, entrapping blue ions until the first gate opens to release them at the opposite side.

gate can comprise any impediment to ion transit, structural studies in primary ion pumps⁸ and cation channels^{9–11} indicate that large domain movements probably control ion transport. But in CLC proteins (channels and pumps), a more subtle structural change — rotation of a single amino-acid side chain — seems to regulate ion passage⁵. This side chain, of a well-conserved glutamate at position 148 (E148), visibly obstructs the extracellular entrance to the ion pathway in the normal CLC-ec1 protein. But if this glutamate is replaced by alanine or glutamine (E148A or E148Q) the side chain swings out of the pore, and is replaced by a Cl^- ion⁵. CLC-0-type Cl^- channels are, in effect, permanently open when they have the same mutations at their corresponding residue (E166). So the negatively charged side chain of the glutamate functions as a gate that blocks ion passage until it is displaced by a permeating Cl^- ion. An enhanced opening of CLC-0 channels is seen at low pH^{5,12}, and this is probably because the protons neutralize the glutamate side chain — the same effect achieved by the mutations.

According to the simple principle outlined above, this conserved glutamate might represent the sole gate sufficient for a CLC-0 channel pore, and one of the two gates needed by a CLC-ec1 exchange pump. Indeed, Lin and Chen¹³ showed that there is no gate at the cytoplasmic end of the CLC-0 channel pore, consistent with it needing no more than one gate. On the other hand, CLC-ec1 exchange pumps are expected to contain a second gate, and the X-ray structures^{3,5} suggest that two residues — serine at position 107 and tyrosine at 445 — contribute to such a gate. The side chains of both interact with the Cl^- ion found in the normal, as well as the E148-mutant, CLC-ec1 near the intracellular end of the route that the Cl^- ions take^{3,5}.

If we pursue this simple analysis, the

external gate ought to be largely absent from the E148A mutant of CLC-ec1. It might therefore be expected to behave like a Cl^- channel, gated only by the intracellular-end gate. But does it? Well, in new reports in this issue and elsewhere^{1,7}, Accardi and Miller show that currents flowing through this CLC-ec1 mutant are perfectly Cl^- selective, just as they are in CLC channels¹. Also, as anticipated, neither currents nor labelled Cl^- fluxes respond to changes in pH^{1,7}, confirming that Cl^- ion movements through this mutant transporter are no longer coupled to proton movements. However, no more than 100,000 Cl^- ions flow, on average, through a mutant E148A CLC-ec1 molecule in a second^{1,7}, whereas Cl^- flow through a single open E166A CLC-0 channel is 100 times faster⁵.

Despite this tiny average current, could the E148A mutant CLC-ec1 exchange pump still behave like a channel? Could, perhaps, the still-remaining internal gate keep the pathway shut 99% of the time, and undergo only brief (as yet undetected) openings that allow Cl^- ion flow through the mutant transporter as rapidly as through an open CLC-0 channel? Probably not, because, at least in the normal exchange pump, both external and internal gates must open and close very frequently to account for the observed rapid exchange of two Cl^- ions for each proton¹. Well, perhaps the internal gate does open and close frequently in the mutant transporter but, when open, Cl^- ion flow is 100-fold slower than through open CLC-0 channels? This doesn't seem likely either, as the equivalent mutation in CLC-0 (E166A) leaves the open-channel current unaltered⁵. More likely, as inferred by the authors¹, discrete movements of Cl^- ions, perhaps still two at a time, 100,000 times each second, persist in the mutant transporter, albeit uncoupled from protons. Clearing this up will require higher resolu-

tion current recordings, as well as information on the whereabouts and workings of the exchange pump's inner gate.

Answers to questions like these, long the stuff of dreams for transport buffs, now seem tantalizingly within reach. There can be no doubt that CLC-ec1 is an exchange pump, just as CLC-0 is undoubtedly a channel. Moreover, a lot of evidence — not just the conserved glutamate gate — makes the structure of CLC-ec1 a reliable template for the architecture of CLC channels^{3,14–16}. The strong implication is that when the structure of the CLC-0 pore is eventually determined it will closely resemble that of CLC-ec1. In the meantime, we can think hard about the functional distinction between a slowly conducting channel and a rapidly gating transporter, when both translocate just one kind of ion. Anyone for ping-pong? ■

David C. Gadsby is in the Laboratory of Cardiac/Membrane Physiology, The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399, USA.

e-mail: gadsby@rockefeller.edu

- Accardi, A. & Miller, C. *Nature* **427**, 803–807 (2004).
- Maduke, M., Pheasant, D. J. & Miller, C. *J. Gen. Physiol.* **114**, 713–722 (1999).
- Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. *Nature* **415**, 287–294 (2002).
- Iyer, R., Iverson, T. M., Accardi, A. & Miller, C. *Nature* **419**, 715–718 (2002).
- Dutzler, R., Campbell, E. B. & MacKinnon, R. *Science* **300**, 108–112 (2003).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. *Nature* **414**, 43–48 (2001).
- Accardi, A., Kolmakova-Partensky, L., Williams, C. & Miller, C. *J. Gen. Physiol.* **123**, 109–119 (2004).
- Toyoshima, C. & Nomura, H. *Nature* **418**, 605–611 (2002).
- Unwin, N., Miyazawa, A., Li, J. & Fujiyoshi, Y. *J. Mol. Biol.* **319**, 1165–1176 (2002).
- Jiang, Y. *et al.* *Nature* **417**, 515–522 (2002).
- Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. *Nature* **423**, 42–48 (2003).
- Chen, M. F. & Chen, T. Y. *J. Gen. Physiol.* **118**, 23–32 (2001).
- Lin, C. W. & Chen, T. Y. *J. Gen. Physiol.* **122**, 147–159 (2003).
- Miller, C. *Phil. Trans. R. Soc. Lond. B* **299**, 401–411 (1982).
- Estévez, R., Schroeder, B. C., Accardi, A., Jentsch, T. J. & Pusch, M. *Neuron* **38**, 47–59 (2003).
- Chen, M. F. & Chen, T. Y. *J. Gen. Physiol.* **122**, 133–145 (2003).

A move to sort life from death

David R. Kaplan and Freda D. Miller

Nerve growth factor determines neuronal cell fate during development or after injury. A newly identified 'death factor', an unprocessed form of this protein, induces cell death by activating two receptors in concert.

The death of nerve cells is a key aspect in the establishment of functional neural circuits during development, but inevitably also features in injury or degenerative conditions. On page 843 of this issue, Nykjaer *et al.*¹ provide some surprising molecular insights into cellular demise. Two unrelated cell-surface proteins, the p75 neurotrophin receptor (p75NTR) and sortilin, collaborate to induce death in responsive cells — including neurons. This joint effort involves direct interactions between p75NTR and sortilin, and is a consequence of the ability of sortilin to bind to an unprocessed form of nerve growth factor (NGF) as well as to the protein neurotensin. p75NTR has been implicated in cell death during both development and injury, but this unexpected partnership has profound consequences for our understanding of cell death in the nervous system, where these receptors cohabit on many of the same cells.

What is surprising about this 'death system' is that NGF was discovered as a factor that promotes the survival and growth of neurons, a finding for which Rita Levi-Montalcini and Stanley Cohen were awarded the Nobel prize in 1986 (ref. 2). Although it binds to p75NTR³, attempts to demonstrate that this receptor mediates NGF's effects largely failed. A second NGF receptor, TrkA, was later shown⁴ to mediate most of the survival and growth effects of NGF (Fig. 1a). So p75NTR, as it physically interacts with TrkA in some circumstances, was largely thought to have some sort of positive accessory role.

A breakthrough came when p75NTR was shown to be essential for the death of several

cell types, including developing neurons^{4,5}. But how could NGF promote survival through TrkA and death through p75NTR — sometimes in the same cell type? It soon became clear that the function of p75NTR depends on the cellular context; activation of p75NTR promotes death in numerous cells, including injured neurons, but promotes migration, growth and survival in other cells⁴.

In the next development, NGF was found to exist in both unprocessed ('pro') and mature forms. On some cells the mature NGF preferentially activates TrkA, whereas proNGF only activates p75NTR⁵. Importantly, proNGF is much more efficient than NGF at inducing the death of responsive cells, leading to speculation that the nature of the NGF itself is a key determinant of the ultimate outcome of p75NTR activation.

It is within this context that the paper of Nykjaer *et al.*¹ now appears. Their results provide a molecular explanation for why p75NTR signals the downfall of some cell types but not others. First, they found that sortilin, a protein that shuttles other proteins within cells and acts as a receptor for neurotensin⁶, interacted with proNGF. Specifically, sortilin bound the 'pro' region of NGF, which is cleaved off during the formation of mature NGF, whereas p75NTR bound regions in mature NGF. Together, a sortilin–p75NTR complex bound proNGF (Fig. 1b). But preventing proNGF binding to sortilin blocked formation of this complex and stopped proNGF binding to p75NTR with high affinity. Such high-affinity interactions occur at the very low concentrations at which proNGF induces cell death.

Nykjaer and colleagues¹ then examined whether sortilin was essential for proNGF to induce cell death through p75NTR. Inhibiting sortilin expression in neurons that express p75NTR, and that normally die when exposed to proNGF, prevented cell death. So what happened when sortilin was put into cells that don't normally have it? They died in response to proNGF. Thus, the presence or absence of sortilin determines whether or not p75NTR functions as a death receptor within a given cell type (Fig. 1).

These findings therefore represent a considerable advance in our understanding of cell death in the nervous system, and clarify many of the seemingly contradictory results obtained with p75NTR in different cell types. Sortilin is broadly expressed in the nervous system⁷, but p75NTR is present during development and when cells of many types are injured³. So this co-receptor system could have a major influence on whether neurons live or die.

Nykjaer and colleagues' study typically raises as many questions as it answers. First, sortilin is required for p75NTR-induced cell death in cultured cells, but does it function similarly in the whole organism? In fact, p75NTR activation does not inevitably cause death, even in cells expressing sortilin⁴. Is this because unprocessed forms of other NGF family members are not synthesized or secreted in their vicinity and/or because all p75NTR in these cells is bound to other receptors? In this regard, p75NTR binds at least two other receptors, TrkA and the Nogo receptor^{3,8,9}. The latter binds proteins in myelin (which ensheathes neurons) that inhibit the regeneration of neuronal processes⁹. The interaction between p75NTR and TrkA enhances NGF binding to TrkA (Fig. 1a), and the interaction between p75NTR and the Nogo receptor is required for growth inhibition^{3,8,9} (Fig. 1c). So, the various and contradictory effects of p75NTR could come down to its choice of co-receptor in a particular cell, or even regions of those cells.

Second, the death signal that emanates from the p75NTR–sortilin complex still eludes us. As sortilin was first found to regulate protein movement within cells⁶, perhaps it chaperones p75NTR to specific compartments in cells for a rendezvous with death-inducing proteins. Third, little is known about how the expression of proNGF and neurotensin is regulated during development and injury. Nykjaer *et al.*¹ show here that neurotensin prevents proNGF's actions, so the relative ratio of these two proteins can determine whether the p75NTR–sortilin complex induces cell death. Answering these questions will be essential for improving our understanding of the basic biology of the nervous system. It will also be necessary for designing effective treatments for the devastating neurological conditions that cause neuronal death.

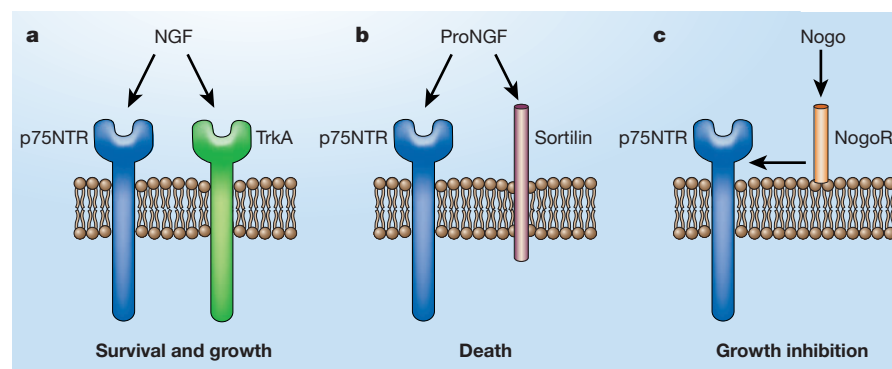


Figure 1 Roles of p75NTR and its co-receptors — TrkA, sortilin and NogoR — in the nervous system. a, The interaction of p75NTR with TrkA enhances the responsiveness of TrkA to nerve growth factor (NGF) to promote survival and growth. b, In contrast, when bound to unprocessed NGF (proNGF), a complex of p75NTR and sortilin induces cell death. c, Nogo causes the complex of p75NTR and the Nogo receptor (NogoR) to inhibit nerve growth.

David R. Kaplan is in the Cancer Research Program, and Freda D. Miller is in the Developmental Biology Program, Hospital for Sick Children, and the Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario M5G 1X8, Canada.
e-mails: dkaplan@sickkids.ca
fredam@sickkids.ca

Superconductivity

Symmetry not required

S. S. Saxena and P. Monthoux

Superconductivity is a complex phenomenon. And now there's something else to think about: a magnetic material whose structure is not mirror symmetric and yet, unexpectedly, superconducts.

Symmetry is central to our understanding and description of natural phenomena. The fundamental conservation laws of physics, such as the conservation of momentum and energy, are the consequences of symmetries in space and time; also, our understanding of the forces of nature is based on a local symmetry known as gauge symmetry. However, the world we observe is often unsymmetrical: in the process of nuclear β -decay, there is an absence of inversion or mirror symmetry, known as parity violation; the structure of DNA is not mirror symmetric.

The consequences of broken symmetries can be dramatic. The breaking of gauge invariance, for example, is associated with the onset of superconductivity — the resistanceless flow of current, usually at very low temperatures. In *Physical Review Letters*, Bauer *et al.*¹ present a material in which space and time inversion symmetries and gauge symmetry are broken: a compound of cerium, platinum and silicon (CePt_3Si) is the first example of a magnetic superconductor that has no mirror symmetry, an observation that will lead to a re-examination of our current understanding of these phenomena.

In 1957, Bardeen, Cooper and Schrieffer (BCS) explained the origin of superconductivity in simple metals such as aluminium. In BCS theory, electrons joined into Cooper pairs are the mediators of the supercurrent flow. The quantum wavefunction (or description) of the superconductor is a coherent superposition of paired-electron states. The Cooper states of lowest energy have overall zero momentum and so there must be pairing of electrons of momenta equal in magnitude, but opposite in direction (Fig. 1a). Moreover, the pair states depend not only on the coordinates of the electrons but also on their spin orientation. Because the electrons in the Cooper-pair state are fermions (with one half-unit of spin each), the pair-state wavefunction must be anti-symmetric under interchange of the

1. Nykjaer, A. *et al.* *Nature* **427**, 843–848 (2004).
2. Levi-Montalcini, R. *Science* **237**, 1154–1162 (1987).
3. Hempstead, B. L. *Curr. Opin. Neurobiol.* **12**, 260–267 (2002).
4. Kaplan, D. R. & Miller, F. D. *Curr. Opin. Neurobiol.* **10**, 381–391 (2000).
5. Lee, R. *et al.* *Science* **294**, 1945–1948 (2001).
6. Mazella, J. *Cell Signal.* **13**, 1–6 (2001).
7. Sarret, P. *et al.* *J. Comp. Neurol.* **461**, 483–505 (2003).
8. Kaplan, D. R. & Miller, F. D. *Nature Neurosci.* **6**, 435–436 (2003).
9. McKerracher, L. & Winton, M. J. *Neuron* **36**, 345–348 (2002).

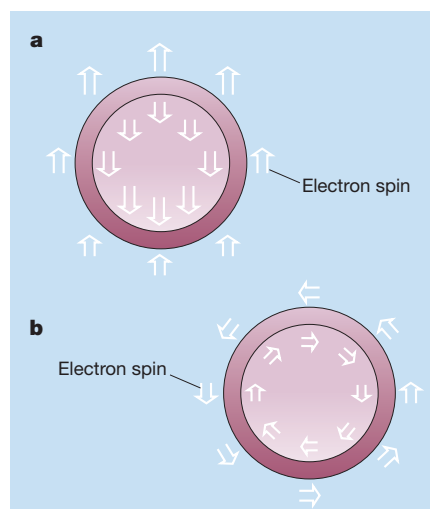


Figure 1 Superconductivity without space inversion symmetry. a, The inner and outer rings (separated for clarity) enclose occupied electron states in momentum space, typical of a normal superconductor. b, When space inversion symmetry is lacking, spins might rotate around the momentum-space surface, clockwise on one side and anticlockwise on the other.

two electrons. This is the origin of the familiar Pauli exclusion principle.

In simple metals such as aluminium, the Cooper-state wavefunction can be separated into two parts, one that depends only on the spatial coordinates and the other only on the spin coordinates. In crystals with space-inversion symmetry, an interchange of the spatial coordinates of the electrons leads to a state that is indistinguishable from the original. Hence, the probability distribution of the electrons in the Cooper state must be unchanged. The Cooper-state wavefunction can only change by a phase factor under this transformation. Because a second interchange must lead back to the original quantum state, the phase factor must be $+1$ or -1 . This simply means that the spatial part of the Cooper state must be either symmetric or anti-symmetric.

For the complete Cooper state to be anti-symmetric, the other part of the wavefunction, the spin part, must have opposite symmetry under electron interchange. Thus, when the spatial wavefunction is symmetric, the spin wavefunction must be anti-symmetric. This arrangement (corresponding to what is known as the 'spin-singlet' state) is realized in many metals, such as aluminium. On the other hand, when the spatial wavefunction is anti-symmetric, the spin wavefunction must be symmetric (corresponding to the 'spin-triplet' state). This particular case is realized in liquid helium-3, which has a superfluid state with intricate properties that arise from the orbital and spin angular momentum of the Cooper-pair states.

Now CePt_3Si , as Bauer *et al.*¹ report, has no space inversion symmetry (Fig. 1), so the spatial part of the wavefunction cannot simply be taken to be symmetric or anti-symmetric. The same applies to the spin part of the wavefunction, if the overall anti-symmetry of the Cooper-pair state is to be preserved. In other words, the Cooper pair in crystals like CePt_3Si is a mixture of spin-singlet and spin-triplet states.

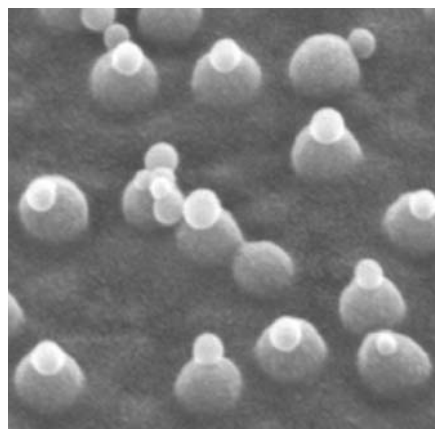
A new way of thinking about the spin configuration of the possible Cooper-pair states for such a parity-violating superconductor is needed (Fig. 1b). In contrast to the previous example of aluminium, where the spin orientation of the fermion is the same over the whole Fermi surface, in this model for a system with broken inversion symmetry the spin orientation would rotate around the surface. The consequences of such a state are not yet fully understood but have been examined in several recent papers^{2–4}.

To progress, it would be desirable to have many examples of such parity-violating superconductors. So it is encouraging that a second example has already been found: Akazawa *et al.*⁵ have reported that the magnetic compound UIr, which also lacks inversion symmetry, becomes superconducting under pressure. Although magnetically mediated pairing is thought to be relevant in such systems, it has thus far only been studied in detail for parity-conserving materials. The discoveries of superconductivity in CePt_3Si and UIr promise to open up entirely new avenues for theoretical and experimental research, and not only in the field of superconductivity.

S. S. Saxena is at the Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.
e-mail: sss21@cus.cam.ac.uk

P. Monthoux is in the School of Physics and Astronomy, and at the Centre for Science at Extreme Conditions, University of Edinburgh, Edinburgh EH9 3JZ, UK.

1. Bauer, E. *et al.* *Phys. Rev. Lett.* **92**, 027003 (2004).
2. Gor'kov, L. P. & Rashba, E. I. *Phys. Rev. Lett.* **87**, 037004 (2001).
3. Frigeri, P. *et al.* <http://arxiv.org/abs/cond-mat/0311354> (2003).
4. Samokhin, K. V., Zijlstra, E. S. & Bose, S. K. <http://arxiv.org/abs/cond-mat/0311181> (2003).
5. Akazawa, T. *et al.* *J. Phys. Condens. Matter* **16**, L29–L32 (2004).



Surface coatings

Nanoscale snowmen look blank

Adv. Mater. 16, 274–277 (2004)

Moths' eyes don't show up in car headlights as they have an anti-reflective 'stealth' surface, which helps moths evade predators. Some artificial anti-reflective coatings mimic the microscopic structure of moth eyes, which are covered with an array of protuberances around 200 nm high. The ideal shape of these bumps is conical, as this produces a gradual change in refractive index that minimizes reflection.

Similar surface-relief patterns can be made using lithographic etching, but Hye Young Koo and colleagues have devised a cheap 'wet' chemical procedure that approximates the conical profiles by generating snowman-like structures in which a small spherical particle sits atop a larger one (pictured above). Both spheres are polystyrene, with diameters of 50 and 100 nm and a negative surface charge.

The nano-snowmen are made through layer-by-layer self-assembly. First, the substrate is coated with a positively charged polymer, which secures the first layer of spheres. Then the tops of the spheres are 'inked' with the positive polymer using a rubber stamp, enabling the second layer of smaller spheres to be attached. The resulting transparent film has about 2% reflectance: about twice that of commercial plastic coatings with conical protrusions made by a more complex moulding method. **Philip Ball**

Medical technology

What a contrast!

Phys. Med. Biol. 49, 501–508 (2004)

To distinguish between structures in the body, many X-ray techniques rely on the use of a contrast agent, the injection of which is a surgical procedure carrying a small risk. As an alternative, methods are being developed that use coherent X-rays alone to create a contrast between tissues with different refractive indices.

Y. Hwu and colleagues have now shown that such a 'phase-contrast' technique can image blood vessels in real time and at resolutions of less than ten micrometres — a level of detail not seen before, say the authors, even with contrast agents. The team used a beam of 'white' X-rays (containing a range of wavelengths), generated by a synchrotron accelerator. Their images of rats show blood vessels around the heart and eyes in very fine detail.

At present the radiation dose exceeds recommended limits for humans, but Hwu *et al.* claim that the exposure time of about 1 millisecond could be reduced. There are, however, only a handful of specialized medical synchrotrons in the world. So, for the short term at least, the technique will be useful in fundamental studies of blood-vessel development and tumour formation, rather than in treating patients. **Mark Peplow**

Atmospheric chemistry

Flying detectives

J. Geophys. Res. doi:10.1029/2003JD003811 (2004)

For years scientists have simply had to assume that one of the intermediates in polar ozone destruction really exists in the atmosphere. But R. M. Stimpfle *et al.* at last report measurements of the chemical species concerned, the chlorine monoxide dimer CLOOCl, in the stratosphere over the Arctic.

The dimer had been identified in laboratory studies, and models of atmospheric ozone loss have relied on its presumed existence. Actually detecting CLOOCl in the atmosphere has proven elusive, however, as the molecule is thought to exist only in the particularly cold winter conditions over the poles. CLOOCl triggers ozone destruction when the absorption of sunlight breaks it into an oxygen molecule and two chlorine atoms: the latter then destroy two ozone molecules, with two chlorine monoxide radicals being produced that can then reform the chlorine monoxide dimer.

The atmospheric measurements of CLOOCl were obtained on a NASA high-altitude research aircraft deployed from Kiruna, Sweden, in the winter of 1999–2000. The results generally confirm accepted understanding of polar chlorine chemistry and will help improve modelling studies of changes in polar ozone. **Juliane Mössinger**

Endocrinology

Amylin for strong bones

J. Cell Biol. doi:10.1083/jcb.200312135 (2004)

As well as insulin, pancreatic β cells secrete a hormone called amylin, which seems to affect bone remodelling. Bones are made and broken down, or resorbed, continuously. Usually this process is in equilibrium,

maintaining total bone mass.

To understand amylin function, Romain Dacquin *et al.* studied amylin-deficient mice. The animals' only abnormality was an increase in bone resorption; bone formation was unaffected. This imbalance resulted in low bone mass, mimicking the human condition osteoporosis, in which patients have brittle bones.

The mice lacking amylin had more osteoclasts, the cells that resorb bone. Adding amylin to bone cells in culture inhibited the production of osteoclasts, via a pathway that uses an enzyme called ERK1/2 (extracellular signal-regulated protein kinase 1/2). Amylin inhibits the process by which osteoclasts are formed, Dacquin *et al.* conclude, so when the hormone is in short supply, osteoclasts proliferate, engulfing more bone. **Laura Nelson**



Vision

Bird's-eye view

Naturwissenschaften 91, 26–29 (2004)

Oilbirds (*Steatornis caripensis*, above) are nocturnal, fruit-eating birds that live in the tropical rainforests of South America. They have a wing-span of around 1 m and, like some bats, they dwell in caves, forage for food at night and use echolocation to help find their way around. But they also have extraordinarily sensitive eyes.

The birds have a relatively large pupil, report Graham Martin *et al.*, which helps the oilbird eye to collect four times more light than the human eye. The light-sensitive rod cells within the eye are stacked three deep, with a density of about 1,000,000 per mm² — more than double the number usually found in vertebrate eyes. The tiered strategy, which has previously been found only in deep-sea fish, maximizes the chance that every photon of light entering the eye will be intercepted.

But extreme sensitivity comes at a cost.

The retina's tiered structure makes it difficult for the brain to work out exactly where the light has come from, so oilbirds have a poor eye for detail. Martin and colleagues therefore suggest that these nocturnal birds rely on a combination of information from smell and echolocation, as well as from sight, to forage successfully. **Helen R. Pilcher**

RNA editing and death of motor neurons

There is a glutamate-receptor defect in patients with amyotrophic lateral sclerosis.

The aetiology of sporadic amyotrophic lateral sclerosis (ALS), a fatal paralytic disease, is largely unknown. Here we show that there is a defect in the editing of the messenger RNA encoding the GluR2 subunit of glutamate AMPA receptors in the spinal motor neurons of individuals affected by ALS. This failure to swap an arginine for a glutamine residue at a crucial site in the subunit, which occurs normally in the affected brain areas of patients with other neurodegenerative diseases, will interfere with the correct functioning of the glutamate receptors and may be a contributory cause of neuronal death in ALS patients.

In RNA editing, gene-specified codons are altered by a post-transcriptional modification of the base sequence of mRNA. The change from glutamine (Q) to arginine (R) at the Q/R site in the putative second membrane domain of GluR2 results from RNA editing and affects the properties of the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazolepropionate) receptors, for example by decreasing their permeability to calcium ions¹.

Editing at the GluR2 Q/R site in neurons occurs with virtually 100% efficiency throughout life from the embryonic stage. The premature demise of mice that have defective GluR2 mRNA editing is caused by neuronal death² but can be rescued by restoring the RNA-editing function³, indicating that this GluR2 modification is crucial for neuronal survival.

We extracted RNA from single motor neurons isolated with a laser microdissector⁴ from five individuals with ALS and from five normal control subjects (see supplementary information). (Written informed consent was obtained from all subjects or from the bereaved family; the Ethics Committee of the University of Tokyo approved the experimental procedures.)

Editing efficiency was calculated by measuring the difference in digestion patterns of nested GluR2 mRNA products from the polymerase chain reaction with reverse transcription, obtained by using the restriction enzyme *BbvI*, which cuts only the edited mRNA^{5,6} (see supplementary information). The editing efficiency in cerebellar Purkinje cells was also quantified for individuals with ALS or with dentatorubral-pallidoluysian atrophy (DRPLA), and the results compared with those from normal subjects.

The frequency of GluR2 mRNA positivity was not significantly different between the ALS and the control groups (two-sample test for equality of proportions, $P > 0.05$).

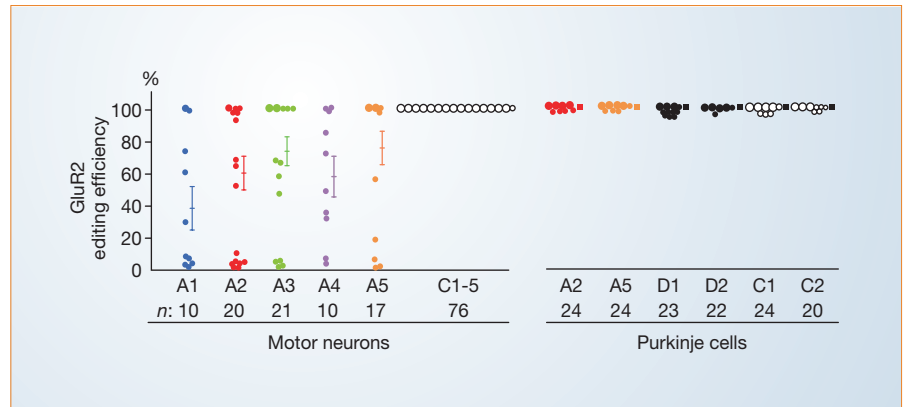


Figure 1 Editing efficiency at the Q/R site of GluR2 messenger RNA in single neurons. Small circles represent the editing efficiency of one cell; each large circle represents five cells in which editing was complete. The mean $\pm$ s.e.m. and the number (n) of cells examined per person are shown. Left, GluR2 mRNA editing is significantly decreased in motor neurons from all patients with ALS (A1–A5) and showed high variability even within individuals (from $75.3 \pm 9.9\%$ in A5 to $38.1 \pm 13.1\%$ in A1), whereas editing was complete in all control motor neurons (C1, $n=28$; C2, $n=12$; C3, $n=13$; C4, $n=12$; C5, $n=11$) (Mann–Whitney U -test, $P < 0.001$). Right, editing efficiency in Purkinje cells from individuals with ALS (A2, $99.9 \pm 0.04\%$; A5, $99.9 \pm 0.06\%$) and individuals with DRPLA (D1, $98.8 \pm 0.5\%$; D2, $99.8 \pm 0.2\%$) was the same as that of control cells (C1, $99.8 \pm 0.1\%$; C2, $99.9 \pm 0.05\%$) ($P > 0.05$).

However, the editing efficiency varied between 0% and 100% in the motor neurons from each individual with ALS, and was incomplete in 44 of them (56%); all 76 of the control motor neurons examined showed 100% editing efficiency (Fig. 1).

The editing efficiency in Purkinje cells was virtually complete in the ALS, DRPLA and normal groups (Fig. 1). Other factors that might influence the properties of AMPA receptors, including the absolute amount of expression and the relative proportion of GluR2 mRNA to total AMPA receptor mRNA, were the same for ALS motor neurons and control motor neurons⁴.

In agreement with our results, mice transgenic for GluR2 made artificially permeable to calcium ions develop motor-neuron disease late in life⁷, indicating that motor neurons may be specifically vulnerable to defective RNA editing. An unedited GluR2 subunit is incorporated into functional AMPA receptors and transported to the cell surface more efficiently than an edited GluR2 subunit⁸, implying that a moderate reduction in GluR2 RNA editing may induce an ALS-like syndrome.

At the tissue level, GluR2 Q/R-site editing is preserved in the severely pathologically affected brain areas of patients with other neurodegenerative diseases⁹. In addition, GluR2 editing is virtually complete in Purkinje cells (Fig. 1) and in the motor cortex⁵ of ALS patients, indicating that the defect in GluR2 editing is likely to be specific to ALS spinal motor neurons.

Further investigation should determine the molecular mechanism underlying the

degeneration of upper motor neurons. If the marked reduction that we observe in GluR2 RNA editing at the Q/R site is relevant to ALS aetiology, elucidation of this mechanism, including the dysfunction of the RNA-editing enzymes involved, could reveal a therapeutic target that is specific to ALS, which may turn out to be a disease of RNA processing¹⁰.

Yukio Kawahara*, Kyoko Ito*, Hui Sun*, Hitoshi Aizawa†, Ichiro Kanazawa*‡, Shin Kwak*

*Department of Neurology, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan
e-mail: kwak-ty@umin.ac.jp

†First Department of Internal Medicine, Asahikawa Medical College, 1-1-1 Midorigaoka Higashinijyo, Asahikawa-shi, Hokkaido 078-8510, Japan

‡National Center of Neurology and Psychiatry, SORST, Japan Science and Technology Corporation, 4-1-1 Ogawa-Higashi, Kodaira-shi, Tokyo 187-8502, Japan

1. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. *Cell* **67**, 11–19 (1991).
2. Brusa, R. *et al. Science* **270**, 1677–1680 (1995).
3. Higuchi, M. *et al. Nature* **406**, 78–81 (2000).
4. Kawahara, Y. *et al. J. Neurochem.* **85**, 680–689 (2003).
5. Takuma, H., Kwak, S., Yoshizawa, T. & Kanazawa, I. *Ann. Neurol.* **46**, 806–815 (1999).
6. Kawahara, Y., Ito, K., Sun, H., Kanazawa, I. & Kwak, S. *Eur. J. Neurosci.* **18**, 23–33 (2003).
7. Feldmeyer, D. *et al. Nature Neurosci.* **2**, 57–64 (1999).
8. Greger, I. H., Khatri, L., Kong, X. & Ziff, E. B. *Neuron* **40**, 763–774 (2003).
9. Akbarian, S., Smith, M. & Jones, E. *Brain Res.* **699**, 297–304 (1995).
10. Lin, C.-L. G. *et al. Neuron* **20**, 589–602 (1998).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Magnetic enhancement of superconductivity

An inhomogeneous superconducting state, not yet conclusively identified, was predicted by Fulde and Ferrell¹ and Larkin and Ovchinnikov² (FFLO) to arise in superconductors with strong Pauli limiting, a consequence of the electrons' Zeeman (spin) energy in a magnetic field. Radovan *et al.*³ propose that the observed cascades of steps in magnetization of the heavy fermion superconductor CeCoIn₅, within the recently discovered second low-temperature state^{3,4}, are due to transitions between Landau-level (LL) states with different *m*-quanta vortices, expected under certain conditions when the magnetic field is swept within the FFLO state^{5–7}. The authors then conclude that the observed steps in magnetization constitute a proof that the low-temperature state in CeCoIn₅ is indeed an FFLO state. We argue that this interpretation of the observed steps in magnetization cannot be supported on either quantitative or qualitative grounds.

The relative effectiveness of the orbital⁸ and paramagnetic⁹ (Pauli) limiting effects in suppressing superconductivity is reflected in the Maki parameter $\alpha = \sqrt{2H_{c2}^0/H_p}$. A large α indicates a small Pauli limiting field, H_p (stronger Pauli limiting), and/or a large orbital limiting field, H_{c2}^0 . Radovan *et al.*³ obtain $\alpha \approx 13$ by using $H_p \approx 4$ T, which assumes an electron *g*-factor of 2 for CeCoIn₅. But the experimental value of the superconducting critical field is $H_{c2} = 12$ T, three times the Pauli (upper) limit used by Radovan *et al.* Therefore, $H_p = 4$ T is unphysical.

H_p can be estimated both theoretically and experimentally. It has been estimated by fitting the critical field H_{c2} of CeCoIn₅ to a model for a *d*-wave superconductor with an FFLO state¹⁰ to give an electron *g*-factor of 0.64 or $H_p = 12.8$ T. A very conservative lower bound of $H_p > 10$ T can be made simply by noting that $H_{c2} = 10$ T at 1 K, where an FFLO state cannot influence H_{c2} .

Theoretically, the Maki parameter must exceed 9 for an *m* ≠ 0 LL state to become a ground state⁵. $H_p = 12.8$ T gives $\alpha \approx 4.5$, which is a third of the value derived by Radovan *et al.*, and half the minimum required for observation of *m* ≠ 0 LL superconducting states. The number of observed steps (tens) also seems to be orders of magnitude greater than that expected theoretically⁵ for a material with α close to that of CeCoIn₅. Only one non-zero state (*m* = 1) is expected⁵ for $\alpha \geq 9$, and two such states (*m* = 1, 2) appear⁵ at $\alpha \geq 20$, far greater than $\alpha \approx 4.5$ for CeCoIn₅.

In addition, the number of higher LL states is theoretically expected to increase as the applied field approaches the parallel

orientation ($\theta \rightarrow 0$), with $m \rightarrow \infty$ at $\theta = 0$ (refs 6, 7). This prediction opposes the trend shown by the results of Radovan *et al.* Thus, we believe that the observed steps in magnetization cannot be due to the higher LL states induced by the FFLO effects in CeCoIn₅.

The heat-capacity data shown in Fig. 2 of Radovan *et al.*³ are inconsistent with our data⁴, and miss a large narrow peak associated with the first-order nature of the superconducting phase transition below 1 K. A related point is that it is also incorrect to identify the order of the phase transition on the basis of the presence or absence of a temperature swing at a phase transition in the described magnetocaloric setup (inset to their Fig. 2). Such swings are expected both at first- and at second-order phase transitions¹¹.

Roman Movshovich*, Andrea Bianchi†, Cigdem Capan*, Marcelo Jaime*, R. G. Goodrich‡

*Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

e-mail: roman@lanl.gov

†Hochfeld Labor Dresden, Forschungszentrum Rossendorf, Postfach 51 01 19, 01314 Dresden, Germany

‡Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803-4001, USA

1. Fulde, P. & Ferrell, R. A. *Phys. Rev.* **135**, A550–A563 (1964).

2. Larkin, A. I. & Ovchinnikov, Y. N. *Zh. Eksp. Teor. Fiz.* **47**, 1136 (1964).

3. Radovan, H. *et al. Nature* **425**, 51–55 (2003).

4. Bianchi, A. *et al. Phys. Rev. Lett.* **91**, 187004 (2003).

5. Buzdin, A. I. & Brison, J. P. *Phys. Lett. A* **218**, 359–366 (1996).

6. Buzdin, A. I. & Brison, J. P. *Europhys. Lett.* **35**, 707–712 (1996).

7. Shimahara, H. & Rainer, D. *J. Phys. Soc. Jpn* **66**, 3591–3599 (1997).

8. Ginzburg, V. L. *Sov. Phys. JETP* **4**, 153–160 (1957).

9. Clogston, M. *Phys. Rev. Lett.* **9**, 266–267 (1962).

10. Won, H. *et al. Preprint at* <http://xxx.lanl.gov/cond-mat/0306548> (2003).

11. Jaime, M. *et al. Phys. Rev. Lett.* **89**, 287201 (2002).

Radovan *et al. reply* — Movshovich *et al.* suggest that our calorimetric observation of the FFLO superconducting phase boundary¹ and its angular dependence are inconsistent with their own data reporting the possible presence of an FFLO phase, and that we have incorrectly used magnetothermal measurements to identify the order of the FFLO transition. They also contest that the correlation of the magnetization steps with the calorimetrically identified FFLO superconducting phase of CeCoIn₅ is coincidental. We consider each of these points in turn.

First, our calorimetrically determined FFLO phase boundary is in fact in good agreement with their own phase diagram² for *H*||[110] and with their earlier calorimetric data³ reporting the absence of an FFLO phase transition for *H*||[001] down to 0.1 K. In addition, our measurements of the variation of C_p with field — data used to determine the location of the FFLO and H_{c2} superconducting phase boundaries — agree with new measurements⁴ at 0.5 K for *H*||[110] and with earlier work⁵ at 0.25 K and 0.8 K for

H||[001]. Because our specific-heat measurements only extend down to 0.15 K, we are unable to address the merits of their revised report² of a possible FFLO phase transition at 0.13 K for *H*||[001].

Second, the determination of the second-order nature of the FFLO phase boundary from the absence of a detectable heat pulse in the magnetothermal data presented in ref. 1 for 0.25 K is also in good agreement with the results in ref. 2. As the first-order nature of the H_{c2} transition at that temperature has already been established^{3,6}, the temperature changes at the transition can be identified as the release and absorption of latent heat. Although our determination of a change in the order of H_{c2} at $T_0 = 1.25 \pm 0.15$ K for *H*||[110] contradicts the originally inferred value⁶ of 0.7 K, it is in increasingly good agreement with more recent values⁷ of 1.0 K for *H*||[100] and 1.1 K for *H*||[110].

Third, decreasing the angle, θ , between the applied field and the superconducting layers at constant external field reduces the value of the field component normal to the layers, $B_{\perp}$. This in turn decreases the Landau-level spacing and allows the order parameter to assume higher states, *m*, for a given FFLO wavevector, $Q = Q(B)$. In our case, however, the data were obtained by increasing the field at constant angle. If the total induction remained constant, the Landau index would then decrease in an increasing field. However, the FFLO wavevector also changes with each step, precluding prediction of the optimal Landau index at a given field at the present time.

A full treatment would require consideration of large effective masses, spin-orbit coupling, *g*-values⁷ and the spatial variation in the magnetic flux (that is, finite Ginzburg–Landau parameter, κ). To our knowledge, the first work implementing some of these effects is given in refs 8, 9, but CeCoIn₅ displays a first-order rather than a second-order phase transition at the critical field, and it is not clear how this affects these predictions for the Landau-level quantization. We have already proposed a method of incorporating strong electron–electron coupling and many-body effects to account for the experimentally determined Pauli limit¹.

H. A. Radovan*, N. A. Fortune, T. P. Murphy, S. T. Hannahs, E. C. Palm, S. W. Tozer, D. Hall

*NHMFL, Florida State University, Tallahassee, Florida 32310, USA

e-mail: radovan@magnet.fsu.edu

1. Radovan, H. *et al. Nature* **425**, 51–55 (2003).

2. Bianchi, A. *et al. Phys. Rev. Lett.* **91**, 187004 (2003).

3. Bianchi, A. *et al. Phys. Rev. Lett.* **89**, 137002 (2002).

4. Aoki, H. *et al. J. Phys. Cond. Mat.* **16**, L13–L19 (2004).

5. Ikeda, S. *et al. J. Phys. Soc. Jpn* **70**, 2248–2251 (2001).

6. Tayama, T. *et al. Phys. Rev. B* **65**, 180504 (2002).

7. Won, H. *et al. Preprint at* <http://xxx.lanl.gov/cond-mat/0306548> (2003).

8. Klein, U. *Preprint at* <http://xxx.lanl.gov/cond-mat/0312095> (2003).

9. Klein, U., Rainer, D. & Shimahara, H. *J. Low Temp. Phys.* **118**, 91–104 (2002).

Secondary active transport mediated by a prokaryotic homologue of ClC Cl[−] channels

Alessio Accardi & Christopher Miller

Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA

ClC Cl[−] channels make up a large molecular family, ubiquitous with respect to both organisms and cell types. In eukaryotes, these channels fulfill numerous biological roles requiring gated anion conductance, from regulating skeletal muscle excitability to facilitating endosomal acidification by (H⁺)ATPases. In prokaryotes, ClC functions are unknown except in *Escherichia coli*, where the ClC-ec1 protein promotes H⁺ extrusion activated in the extreme acid-resistance response common to enteric bacteria. Recently, the high-resolution structure of ClC-ec1 was solved by X-ray crystallography. This primal prokaryotic ClC structure has productively guided understanding of gating and anion permeation in the extensively studied eukaryotic ClC channels. We now show that this bacterial homologue is not an ion channel, but rather a H⁺-Cl[−] exchange transporter. As the same molecular architecture can support two fundamentally different transport mechanisms, it seems that the structural boundary separating channels and transporters is not as clear cut as generally thought.

We have recently overcome technical barriers preventing electrophysiological recording of ClC-ec1 and have described conductance properties of this purified protein incorporated into planar lipid bilayer membranes¹. As expected for a ClC channel, this homologue passes Cl[−]-selective current. By ion channel standards, the single-molecule turnover is minuscule (10⁴–10⁵ ions s^{−1}) but macroscopic current arising from hundreds to thousands of ClC-ec1 molecules may readily be observed. The current shows no voltage-dependent gating, but is instead maximally activated by low pH, in accord with its suggested biological function² as a proton-activated electrical 'leak' for Cl[−].

Cl[−] selectivity is imperfect, however, as seen from the 'reversal potential', that is, the voltage needed to null the current in a transmembrane KCl gradient. Currents recorded under salt-gradient conditions (300 mM KCl in the internal solution and 45 mM KCl in the external solution, symmetrical pH 3) in response to a family of voltage pulses are shown in Fig. 1a, along with current–voltage (*I*–*V*) curves derived from these recordings (Fig. 1c). The *I*–*V* curve is displaced along the voltage axis in the positive direction, indicating preferential permeation by Cl[−], but the reversal potential, 30 mV, is substantially below the Nernst equilibrium potential expected for perfect Cl[−] selectivity, 45 mV (arrow). This fact means that some other ion in the system must carry current along with Cl[−].

The missing ion

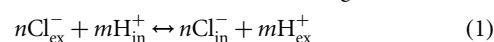
Which of the other ions present accounts for the sub-nernstian reversal potential? We set out to identify this competing ion in our minimal system, where the only ions present, besides Cl[−], are K⁺, H⁺, OH[−] and the buffers glutamate and histidine. K⁺ and the buffers were previously ruled out by ion-substitution experiments¹. Nor can OH[−] explain the discrepancy, as at the low pH of these experiments this anion is present at 10^{−11} M, far too low a concentration to permeate at rates on the order of turnover (10⁵ s^{−1}) given the constraints of diffusion limitation. We are thus left with H⁺ as the only possible candidate. If Cl[−] and H⁺ diffuse through a pore, each under the influence of its own electrochemical gradient, then the reversal potentials obtained above (6.7-fold Cl[−] gradient, symmetrical pH 3) require that H⁺ be 50–100-fold more

permeant than Cl[−]. Because all previously studied ClC channels are specific for small anions, this conclusion is alarming. It is also immediately testable. If pH were to be raised symmetrically, the interfering permeation by H⁺ should be suppressed, and the reversal potential in a Cl[−] gradient should approach the ideal value. But we find that the *I*–*V* curve at pH 5 reverses at precisely the same value as at pH 3 across the entire range of Cl[−] concentrations (Figs 1b, c and 2a). A similar result was found at pH 7 (ref. 1). This is a paradoxical situation, as all ions in the system are ruled out as electrodiffusive interlopers, and yet the sub-nernstian reversal potentials demand that an ion besides Cl[−] carries current.

This other ion is in fact H⁺, as the experiments of Fig. 1d–g show. With symmetrical 300 mM Cl[−], a transmembrane pH gradient (pH 3_{in}/5_{ex}) produces a 35–40-mV shift in the *I*–*V* curve in the direction of the H⁺ equilibrium potential (±117 mV depending on the sidedness of the gradient). Under these symmetrical-salt conditions, the current observed at zero voltage cannot represent diffusive movement of Cl[−]. In terms of pore-mediated ion permeation, these results amplify the paradox above. H⁺ concentration, when varied symmetrically, has no effect on reversal potential, but when varied asymmetrically has a profound influence.

H⁺-Cl[−] countertransport

This conundrum vanishes if we abandon the assumption, natural to make for a Cl[−] channel homologue³, that Cl[−] permeates electrodiffusively through a water-filled pore. If instead we imagine that Cl[−] crosses the membrane in stoichiometric exchange for H⁺, the phenomena above are not only explained, but are required. We propose that ClC-ec1 mediates the Cl[−]-H⁺ exchange reaction



where *n* and *m* are stoichiometric coefficients and subscripts refer to the two sides of the membrane, internal and external. With Cl[−] and H⁺ movements obligatorily coupled, the reversal potential is the voltage at which the reaction is at thermodynamic equilibrium, and may thus be derived without reference to any details of the exchange mechanism

$$V_r = \frac{1}{1+r} [E_{\text{Cl}} + rE_{\text{H}}] \quad (2)$$

where the stoichiometric ratio $r = m/n$, and the Nernst potential of each ion, E_X , is

$$E_X = -\frac{RT}{z_X F} \ln \left(\frac{[X]_{in}}{[X]_{ex}} \right) \quad (3)$$

where z_X is ionic valence, R is universal gas constant, T is temperature, F is Faraday's constant and where brackets denote ionic activity.

This secondary active transport mechanism makes several predictions that are inconsistent with pore electrodiffusion. First, it demands that the reversal potential must vary logarithmically with the concentration ratio of each ion separately, and that the slope of this variation must always be sub-nernstian. Figure 2 illustrates this behaviour with three data sets, two of them varying Cl^- gradient at fixed symmetrical pH (3 or 5), the third varying pH gradient at fixed Cl^- (300 mM); the data adhere well to equation (2), with a global fit determining the stoichiometric ratio, r , as the single adjustable parameter. For the fit shown in Fig. 2a, b, $r = 0.43$; with individual data sets we find this parameter to fall in the range 0.41–0.66, consistent with a stoichiometry of 2 Cl^- ions per H^+ ($r = 0.5$). The slopes of the reversal potential plots are 38 and 18 mV per decade for Cl^- and H^+ , respectively, well below the ideal Nernst slope of 58 mV per decade, whereas the sum of these slopes, 56 mV, is within error identical to the Nernst value, as equation (2) requires. Figure 2a, b also illustrates the futility of forcing an electrodiffusion mechanism, in which Cl^- and H^+ permeate individually through a pore, to agree with the experimental results. Further analysis (not shown) confirms that this inadequacy is a general property of dissipative electrodiffusion models. Also in contrast to pore diffusion, equation (2) predicts that the reversal potential at a fixed Cl^- gradient should remain invariant when pH is changed symmetrically, as observed above (Figs 1c and 2a).

Finally, the H^+ - Cl^- exchange mechanism predicts that the reversal potential measured in the presence of both Cl^- and pH gradients should be the simple sum of those measured with each individual gradient alone (that is, with a Cl^- gradient and symmetrical pH, or vice versa). To test this prediction, we measured reversal potentials with Cl^- and pH gradients simultaneously present (Fig. 2c). These values (filled circles) compare well with

the sum of the experimental values (filled triangles) measured in the individual gradients of Fig. 2a, b. The figure also presents the reversal potentials theoretically expected for the countertransport mechanism with 2 $\text{Cl}^-/1 \text{H}^+$ stoichiometry ($r = 0.5$, equation (2), bold arrows) and for the electrodiffusion model (thin arrows) with H^+/Cl^- permeability ratio determined from the global fit (Fig. 2a, b). In each case, the reversal potentials in the combined gradients are much closer to the exchanger than to the channel predictions, which in some cases get even the sign wrong.

Direct observation of secondary active transport

Despite the solid ground afforded by the thermodynamic considerations above, the idea that ClC-ec1 is not an ion channel is startling enough to call for further experimental confirmation. If this protein operates as a Cl^- - H^+ exchanger, then a pre-established gradient of either ion should drive active transport of its exchange partner. We therefore carried out classical secondary active transport experiments^{4–6} in which a Cl^- gradient was established across liposome membranes and H^+ transport was followed by pH changes in the external solution. In the experiment of Fig. 3a (top trace), ClC-ec1-reconstituted liposomes loaded with 350 mM Cl^- at pH 4.5 were suspended in 3 mM Cl^- at pH 4.8; under these conditions no transport can occur, because the exchanger, being highly electrogenic (three charges moved per turnover), polarizes the membrane. The reaction is initiated by adding the K^+ ionophore valinomycin to allow counterion movement and to hold the liposome voltage at a slightly positive value (5–10 mV). Now, H^+ enters the vesicles against both pH and electrical gradients as Cl^- exits, reaching steady state in about a minute. Addition of the weak-acid proton uncoupler carbonyl cyanide *p*-(trifluoro-methoxy)phenyl hydrazone (FCCP) reverses the H^+ uptake. It is difficult to quantify precisely the electrochemical gradient against which H^+ moves, but an estimate based on the intraliposomal volume and buffer capacity suggests that a steady-state pH gradient of at least 1.5 units is sustained by the transporter. This conservative estimate neglects the additional electrical gradient against which both protons and Cl^- move. It is worth emphasizing that an ion channel permeable to Cl^- and H^+ would produce H^+ efflux under these conditions rather than the influx observed. Additional experiments (data not shown)

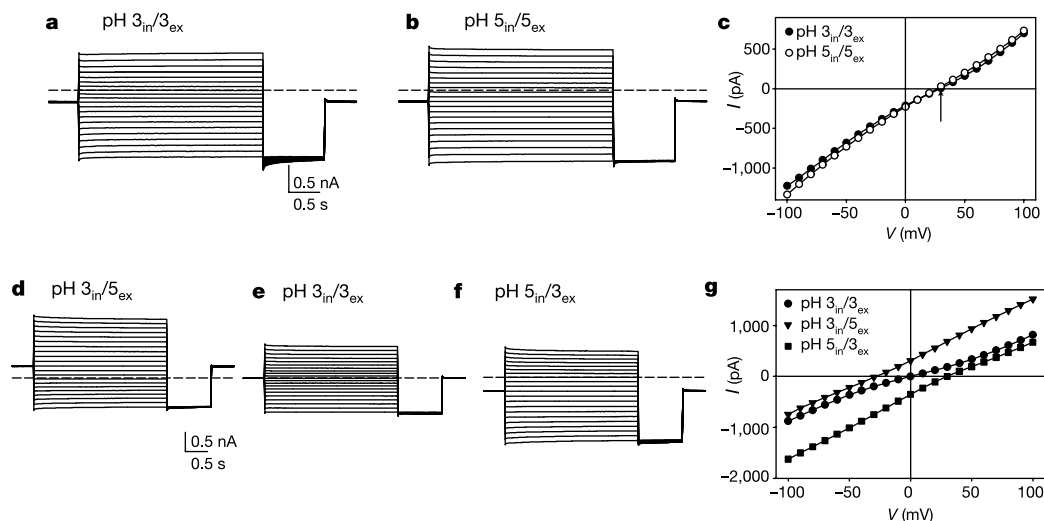


Figure 1 Proton dependence of ClC-ec1 currents. ClC-ec1 currents in response to 3-s voltage pulses from -100 to 100 mV in 10-mV steps followed by a 1-s tail pulse to -100 mV were recorded in various conditions of Cl^- and pH. Holding voltage was 0 mV. **a, b**, KCl gradient of 300/45 mM (internal/external), symmetrical pH 3 (**a**) or pH 5 (**b**). **c**, I - V curves from the traces at pH 3 (filled circles) and pH 5 (open circles). An arrow

marks the Nernst potential for Cl^- . **d**, Symmetrical HS solution (300 mM Cl^-), pH 3_{in}/5_{ex} gradient. **e**, Same bilayer as in **d** but with symmetrical pH 3. **f**, Same bilayer as in **d, e**, with a 5_{in}/3_{ex} pH gradient. **g**, I - V curves from the traces shown in **d-f**. Circles, pH 3_{in}/3_{ex}; squares, pH 5_{in}/3_{ex}; triangles, pH 3_{in}/5_{ex}. Dashed lines represent zero current level.

establish that with an inwardly oriented Cl^- gradient, H^+ is extruded against a pH gradient in a FCCP-sensitive manner, and that no H^+ uptake occurs if CIC-ec1 is omitted from the reconstitution mix.

An exchange mechanism requires reciprocity in transport behaviour; just as a Cl^- gradient drives active H^+ transport, so should a pH gradient drive uphill Cl^- flux. An example of such a reciprocal experiment is shown in Fig. 3b. Here (top trace), an inwardly directed proton gradient is imposed across the liposome membrane, and uphill Cl^- efflux, initiated by valinomycin, is followed by the de-quenching of a Cl^- -sensitive fluorometric dye trapped within the liposomes. This increase in fluorescence is reversed by collapsing the proton gradient with FCCP. A rough calculation based on the approximate doubling of fluorescence implies that the H^+/Cl^- exchange establishes at least a fivefold Cl^- gradient. No such responses were observed in protein-free liposomes.

Abolition of proton coupling by E148A mutation

In CIC-ec1 a conserved glutamate residue, E148, lies close to a bound Cl^- ion, occluding the extracellular side of the transport pathway^{7,8}. Mutating this glutamate to Ala or Gln maximally activates Cl^- fluxes in reconstituted liposomes, completely abolishes their pH sensitivity, and renders the currents nearly ideally Cl^- -selective¹. These results imply that E148 is specifically required for coupling of protons to Cl^- permeation. Accordingly, we reconstituted CIC-ec1(E148A) into liposomes and found the protein inactive in both of the secondary active-transport assays of Fig. 3 (lower traces). These liposomes are still fully functional in Cl^- - Cl^- exchange¹ and net electrogenic downhill Cl^- movement (data not shown); the mutation's effect is to eliminate H^+ coupling to Cl^- flux. This conclusion is further supported by the electrical behaviour of the E148A mutant in planar bilayers. In marked contrast

with wild-type currents (Fig. 1), the I - V curve of the mutant protein (Fig. 4) is completely unresponsive to pH manipulation, with the reversal potential in symmetrical Cl^- remaining at zero with a 4-unit pH gradient across the bilayer. As this mutation apparently retains the low unitary transport rate of wild-type protein¹, we suggest that uncoupled Cl^- translocation through E148A, although now dissipative, is nonetheless mediated by a cycle of conformational changes, not by pore diffusion. This suggestion must remain tentative, however, as macroscopic electrophysiological properties do not readily distinguish a simple carrier mechanism from a channel.

Conclusions and implications

These experiments taken individually and in aggregate establish that this bacterial CIC homologue is not an ion channel as we had originally imagined³, but instead acts as a H^+/Cl^- exchange transporter with a likely stoichiometric ratio of 2 Cl^-/H^+ . This conclusion forces a reinterpretation of the activation of Cl^- currents by low $\text{pH}^{1,2}$ in terms of proton transport rather than proton-dependent channel gating. These results should not be taken to

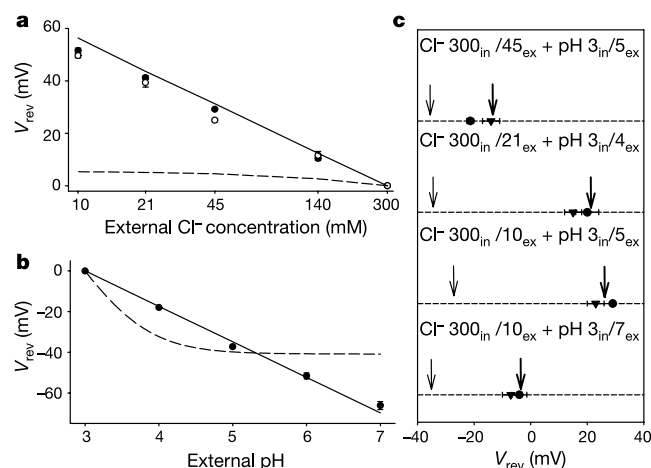


Figure 2 H^+ and Cl^- dependence of reversal potential. CIC-ec1 I - V curves were measured as in Fig. 1, with the internal side solution held fixed at 300 mM Cl^- , pH 3 (HS), while Cl^- and/or pH of the external side was varied. **a**, Reversal potential, V_{rev} , plotted as a function of external Cl^- concentration at symmetrical pH 3 (filled circles) or symmetrical pH 5 (open circles). **b**, V_{rev} plotted as a function of external pH at symmetrical 300 mM Cl^- . Solid lines are global fits on all three data sets for exchange mechanism, equation (2), with $r = 0.43$; dashed lines are global fits for simple electrodiffusion, using the Goldman-Hodgkin-Katz equation³³, with $P_{\text{H}}/P_{\text{Cl}} = 850$. Mean ionic activity coefficients for KCl^{34} were used in both fits. Each point represents mean $\pm$ s.e.m. of 3–8 separate experiments. **c**, Additivity of reversal potentials. Circles, experimental data with combined Cl^- and pH gradients indicated; triangles, sum of the reversal potentials of the appropriate single-gradient data of **a** and **b**; bold arrows, predictions of equation (2) assuming fixed 2:1 Cl^-/H^+ stoichiometry; thin arrows, predictions of electrodiffusion as in **a**, **b**.

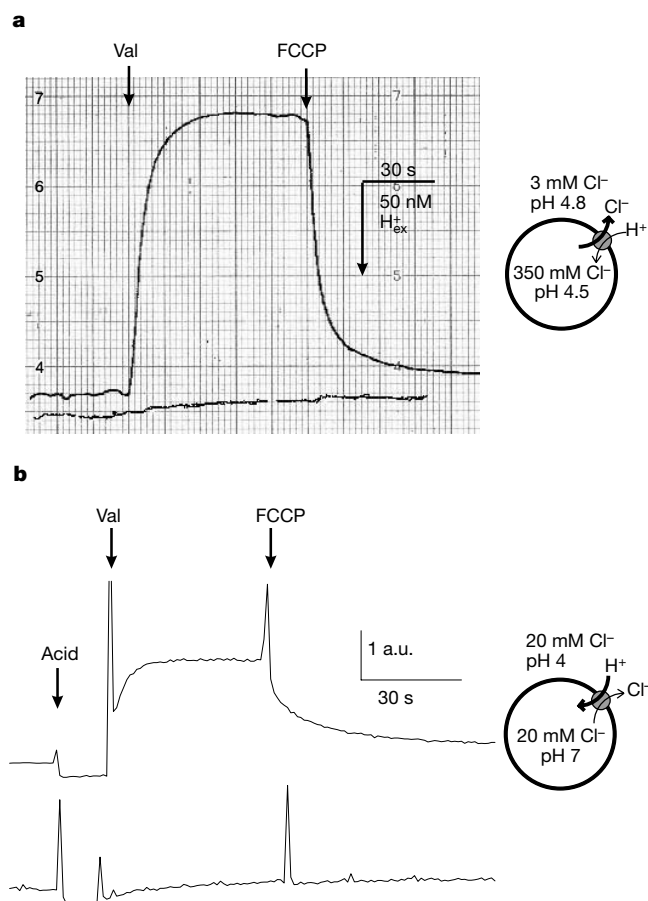


Figure 3 Secondary active transport. **a**, Cl^- -driven H^+ transport. An outwardly directed Cl^- gradient was imposed on CIC-ec1-reconstituted liposomes, and uptake of H^+ was followed by external pH measurement. The uptake corresponds to an increase of external pH of approximately 0.02 pH units. Top trace: wild-type protein; bottom trace, E148A mutant protein. Val, valinomycin. **b**, H^+ -driven Cl^- transport. An inwardly directed H^+ gradient was imposed on reconstituted vesicles, and Cl^- efflux was followed by de-quenching of trapped SPQ. Top trace, wild-type protein; bottom trace, E148A mutant protein. Cartoons represent ionic conditions during transport, bold and thin arrows, respectively, indicate downhill and uphill fluxes. a.u., arbitrary units.

imply that eukaryotic CIC channels are also exchangers; indeed, CIC-0, CIC-1 and CIC-2 are unambiguously Cl^- -selective channels^{9–11}, which, while displaying proton-dependent gating, show no indications of H^+ permeability. The results do raise questions, however, as to the status of other homologues such as CIC-6 and CIC-7, which have thus far been assumed, without proof, to be channels.

Our conclusions are harmonious with the recent X-ray structures of wild-type and mutant CIC-ec1 determined by Dutzler and colleagues^{7,8}. In particular, none of these shows a transmembrane pore, which would be required for a channel but forbidden in any well-coupled exchanger. Figure 5 represents a cut-away view of a single subunit showing the central Cl^- ion (green) and the occluded escape-ways leading to the internal and external solutions. E148 (red) forms a steric barrier separating the Cl^- ion from the external solution, whereas S107 (yellow) and Y445 (blue) obstruct the internal pathway. Because E148 is strongly implicated in linking H^+ and Cl^- transport, these two ions appear to share, at least in part, a common translocation pathway.

For ion channel proteins, a single high-resolution structure—that of the ion-occupied pore—may suffice for illuminating the mechanistic underpinnings of electrodiffusive ion permeation¹². But this is not the case for transporters, as substrate permeation in these proteins arises from a cycle of conformational changes. We suppose that the multiple high-resolution structures of CIC-ec1 (refs 7, 8) include two of the conformations involved in substrate translocation. In the wild-type protein, the buried Cl^- ion is reminiscent of occluded states of P-type ATPase ion pumps^{13,14}. In the E148Q mutant, which presumably models a protonated glutamate, this Cl^- ion becomes exposed to the extracellular solution as the side chain rotates away from the anion to open an external ‘gate’⁸. Is this a protonated outward-facing conformation of the exchange cycle? We cannot answer this question now, nor can we propose a detailed mechanism for substrate transport in terms of the structure; we

suggest, however, that the conformational changes involved in transport are relatively minor, as the unitary turnover rate of 10^5 s^{-1} , although low for a channel, is high for a transporter. Such a picture would differ from that of the MFS transporter family, for which the countertransporters OxlT¹⁵ and GlpT¹⁶ and the co-transporter LacY¹⁷ all show wide aqueous cavities that are thought to alternate in exposure to the two sides of the membrane via large polypeptide backbone rearrangements.

Finally, we are impressed by the success of the bacterial CIC structure in predicting and rationalizing electrophysiological behaviour of eukaryotic CIC channels^{8,18–21}. This is, in retrospect, unexpected, as CIC-0 and CIC-1 are ion channels but CIC-ec1 is not. It is notable that mutations of the conserved glutamate produce constitutive activity and eliminate the normal pH-dependence of both CIC-0 and CIC-ec1. This functional congruence on mechanistically dissimilar backgrounds suggests that structurally subtle alterations accumulated through CIC evolution might precipitate a profound mechanistic switch. Unusual as it is, this situation is not unprecedented. The widespread family of ABC transporters²², whose members operate as ATP-driven pumps—inward for nutrients or outward for xenobiotics—offers the striking example of CFTR, in which the protein, although maintaining molecular hallmarks of the family, functions not as a solute pump but rather as an ATP-gated Cl^- channel²³. Moreover, a P-type ATPase ion pump is known to degrade to an ion channel under the influence of a natural toxin that opens the transport pathway to both sides of the membrane simultaneously²⁴; an analogous ‘slippage’ may also operate in several neurotransmitter-reuptake transporters^{25–27}. Finally, our results bring to mind the puzzling finding that CIC-0 gating exhibits tight coupling to Cl^- permeation, as though this ion channel carries mechanistic remnants of a transporter conformational cycle^{28–30}. These observations collectively suggest that in structural terms, transporters and channels may be separated by an exceedingly fine line. □

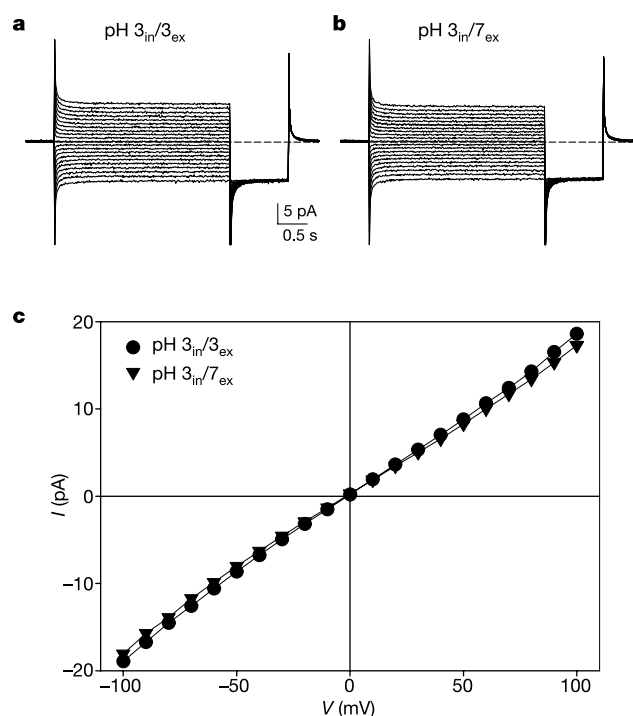


Figure 4 E148A mutation abolishes H^+ dependence of reversal potential. **a, b**, Currents were recorded as in Fig. 1 from the E148A mutant in symmetrical HS solution at pH 3 (**a**) or with a pH 3–7 gradient (**b**). **c**, I - V curves from the traces in **a** and **b**.

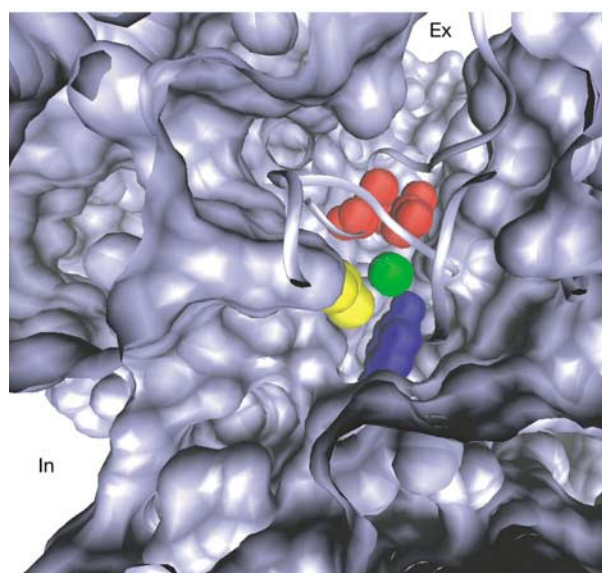


Figure 5 Occlusion of the central Cl^- ion in CIC-ec1. A cut-away view transecting the central Cl^- -coordination region of a single CIC-ec1 subunit is shown in surface representation. Ribbons represent parts of the protein that would otherwise obscure the view of the central Cl^- ion (green) and the extracellular access pathway. The figure emphasizes residues E148 (red), S107 (yellow) and Y445 (blue) that block pathways to extracellular or intracellular solutions. The figure was prepared with the VMD program³⁵.

Methods

Biochemical procedures and solutions

Methods for expression, purification, reconstitution into liposomes, and electrophysiological recording of ClC-ec1 have been described in detail¹. Protein was reconstituted at 5–25 µg mg^{−1} lipid for liposomes used in flux assays, and at 50–75 µg mg^{−1} for liposomes fused into planar lipid bilayers. For bilayer recording, liposomes were formed in 450 mM KCl, 25 mM citric acid, 25 mM phosphoric acid, adjusted to pH 7.0 with KOH. The standard high-salt (HS) solution was 290 mM KCl, 10 mM HCl, 5 mM histidine base, 5 mM glutamic acid, adjusted to the desired pH with KOH; KCl was changed in solutions using lower Cl[−] concentrations. Typically, currents were in the range 0.05–2 nA at 100 mV holding voltage, except for the E148A mutant, for which the currents tend to run down with time.

Coupled fluxes of Cl[−] and H⁺

Uphill movement of Cl[−] driven by a pH gradient was followed by the Cl[−]-sensitive fluorophore 6-methoxy-*N*-(3-sulphopropyl)-quinolinium (SPQ)³¹. SPQ (300 µM) was added to a suspension of liposomes (20 mg ml^{−1}) formed in 20 mM KCl, 20 mM citrate-phosphate, pH 7.0 (RH buffer); the liposomes were frozen at −80 °C and sonicated briefly to give unilamellar vesicles trapping the fluorophore. External SPQ was removed by spinning a 100-µl aliquot through a 1.5-ml Sephadex G-50 column equilibrated in RH buffer, which was immediately added to 1.9 ml of RH buffer in a stirred fluorimeter cuvette. Fluorescence was measured with excitation and emission at 317 and 445 nm, respectively. A pH gradient was set up across the liposome membranes by acidifying the external solution to pH 4.0 with citric acid, and after a stable baseline reading had been confirmed, Cl[−] efflux was initiated by adding valinomycin (1 µg ml^{−1}); after steady-state fluorescence was reached, the H⁺ gradient was collapsed with FCCP (0.25 µg ml^{−1}).

Uphill movement of H⁺ driven by a Cl[−] gradient was followed by recording the pH of a lightly buffered vesicle suspension³². Reconstituted vesicles loaded with high Cl[−] medium (350 mM KCl, 50 mM citrate, 20 mM phosphate, pH 7.0) were acidified to pH 4.5 with phosphoric acid, and were frozen, thawed and sonicated as above. External Cl[−] was lowered by spinning 100 µl aliquots through Sephadex G-50 columns equilibrated 3 mM KCl, 300 mM K₂SO₄, 2 mM glutamic acid, pH 4.8. The resulting spin-through (100 µl) was then diluted into 1.9 ml containing this low-Cl[−] solution in a stirred cell monitored by a recording pH meter. Proton uptake (indicated by an increase in external pH) was initiated by valinomycin as above, and the experiment was terminated by collapsing the pH gradient with FCCP.

Received 14 November 2003; accepted 5 January 2004; doi:10.1038/nature02314.

- Accardi, A., Kolmakova-Partensky, L., Williams, C. & Miller, C. Ionic currents mediated by a prokaryotic homologue of ClC Cl[−] channels. *J. Gen. Physiol.* **123**, 109–119 (2004).
- Iyer, R., Iverson, T. M., Accardi, A. & Miller, C. A biological role for prokaryotic ClC chloride channels. *Nature* **419**, 715–718 (2002).
- Maduke, M., Pheasant, D. J. & Miller, C. High-level expression, functional reconstitution, and quaternary structure of a prokaryotic ClC-type chloride channel. *J. Gen. Physiol.* **114**, 713–722 (1999).
- Foster, D. L., Garcia, M. L., Newman, M. J., Patel, L. & Kaback, H. R. Lactose-proton symport by purified lac carrier protein. *Biochemistry* **21**, 5634–5638 (1982).
- Schellenberg, G. D. & Swanson, P. D. Solubilization and reconstitution of membranes containing the Na⁺-Ca²⁺ exchange carrier from rat brain. *Biochim. Biophys. Acta* **690**, 133–144 (1982).
- Weinman, E. J., Shenolikar, S., Cragoe, E. J. Jr & Dubinsky, W. P. Solubilization and reconstitution of renal brush border Na⁺-H⁺ exchanger. *J. Membr. Biol.* **101**, 1–9 (1988).
- Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
- Dutzler, R., Campbell, E. B. & MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300**, 108–112 (2003).
- Miller, C. Open-state substructure of single chloride channels from *Torpedo* electroplax. *Phil. Trans. R. Soc. B* **299**, 401–411 (1982).
- Pusch, M., Steinmeyer, K. & Jentsch, T. J. Low single channel conductance of the major skeletal muscle

chloride channel, ClC-1. *Biophys. J.* **66**, 149–152 (1994).

- Weinreich, F. & Jentsch, T. J. Pores formed by single subunits in mixed dimers of different ClC chloride channels. *J. Biol. Chem.* **276**, 2347–2353 (2001).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Shani, M., Goldschleger, R. & Karlish, S. J. Rb⁺ occlusion in renal (Na⁺ + K⁺)-ATPase characterized with a simple manual assay. *Biochim. Biophys. Acta* **904**, 13–21 (1987).
- Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, 605–611 (2002).
- Hirai, T. *et al.* Three-dimensional structure of a bacterial oxalate transporter. *Nature Struct. Biol.* **9**, 597–600 (2002).
- Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**, 616–620 (2003).
- Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
- Lin, C. W. & Chen, T. Y. Probing the pore of ClC-0 by substituted cysteine accessibility method using methane thiosulfonate reagents. *J. Gen. Physiol.* **122**, 147–159 (2003).
- Chen, M. F. & Chen, T. Y. Side-chain charge effects and conductance determinants in the pore of ClC-0 chloride channels. *J. Gen. Physiol.* **122**, 133–145 (2003).
- Estevez, R., Schroeder, B. C., Accardi, A., Jentsch, T. J. & Pusch, M. Conservation of chloride channel structure revealed by an inhibitor binding site in ClC-1. *Neuron* **38**, 47–59 (2003).
- Traverso, S., Elia, L. & Pusch, M. Gating competence of constitutively open ClC-0 mutants revealed by the interaction with a small organic inhibitor. *J. Gen. Physiol.* **122**, 295–306 (2003).
- Davidson, A. L. Mechanism of coupling of transport to hydrolysis in bacterial ATP-binding cassette transporters. *J. Bacteriol.* **184**, 1225–1233 (2002).
- Baukowitz, T., Hwang, T.-C., Nairn, A. C. & Gadsby, D. C. Coupling of CFTR Cl[−] channel gating to an ATP hydrolysis cycle. *Neuron* **12**, 473–482 (1994).
- Artigas, P. & Gadsby, D. C. Na⁺/K⁺-pump ligands modulate gating of palytoxin-induced ion channels. *Proc. Natl Acad. Sci. USA* **100**, 501–505 (2003).
- Mager, S. *et al.* Conducting states of a mammalian serotonin transporter. *Neuron* **12**, 845–859 (1994).
- Fairman, W. A., Vandenberg, R. J., Arriza, J. L., Kavanaugh, M. P. & Amara, S. G. An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. *Nature* **375**, 599–603 (1995).
- Adams, S. V. & DeFelice, L. J. Ionic currents in the human serotonin transporter reveal inconsistencies in the alternating access hypothesis. *Biophys. J.* **85**, 1548–1559 (2003).
- Richard, E. A. & Miller, C. Steady-state coupling of ion-channel conformations to a transmembrane ion gradient. *Science* **247**, 1208–1210 (1990).
- Chen, T.-Y. & Miller, C. Nonequilibrium gating and voltage dependence of the ClC-0 Cl[−] channel. *J. Gen. Physiol.* **108**, 237–250 (1996).
- Pusch, M., Ludewig, U., Rehfeldt, A. & Jentsch, T. J. Gating of the voltage-dependent chloride channel ClC-0 by the permeant anion. *Nature* **373**, 527–531 (1995).
- Biwersi, J., Tulk, B. & Verkman, A. S. Long-wavelength chloride-sensitive fluorescent indicators. *Anal. Biochem.* **219**, 139–143 (1994).
- Mitchell, P. & Moyle, J. Stoichiometry of proton translocation through the respiratory chain and adenosine triphosphatase systems of rat liver mitochondria. *Nature* **208**, 147–151 (1965).
- Hille, B. *Ion Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 2001).
- Robinson, R. A. & Stokes, R. H. *Electrolyte Solutions* (Butterworths, London, 1965).
- Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 27–28, 33–38 (1996).

Acknowledgements We are grateful to C. Williams, B. Bablouzion and L. Kolmakova-Partensky for technical assistance, to C. Nimigeon for advice throughout the course of this work, and to M. Walden for help with fluorescence measurements. We also thank P. De Weer, F. M. Ashcroft and G. Yellen for critical readings of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.M. (cmiller@brandeis.edu).

Structure of the signal recognition particle interacting with the elongation-arrested ribosome

Mario Halic¹, Thomas Becker¹, Martin R. Pool², Christian M. T. Spahn³, Robert A. Grassucci⁴, Joachim Frank⁴ & Roland Beckmann¹

¹Institute of Biochemistry, Charité, University Medical School, Humboldt University of Berlin, Monbijoustrasse 2, 10117 Berlin, Germany

²University of Manchester, School of Biological Sciences, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK

³Institute of Medical Physics and Biophysics, Charité, University Medical School, Humboldt University of Berlin, Ziegelstrasse 9, 10117 Berlin, Germany

⁴Howard Hughes Medical Institute, Health Research Incorporated, Wadsworth Center, Empire State Plaza, Albany, New York 12201, and Department of Biomedical Sciences, State University of New York at Albany, Albany, New York 12222, USA

Cotranslational translocation of proteins across or into membranes is a vital process in all kingdoms of life. It requires that the translating ribosome be targeted to the membrane by the signal recognition particle (SRP), an evolutionarily conserved ribonucleoprotein particle. SRP recognizes signal sequences of nascent protein chains emerging from the ribosome. Subsequent binding of SRP leads to a pause in peptide elongation and to the ribosome docking to the membrane-bound SRP receptor. Here we present the structure of a targeting complex consisting of mammalian SRP bound to an active 80S ribosome carrying a signal sequence. This structure, solved to 12 Å by cryo-electron microscopy, enables us to generate a molecular model of SRP in its functional conformation. The model shows how the S domain of SRP contacts the large ribosomal subunit at the nascent chain exit site to bind the signal sequence, and that the Alu domain reaches into the elongation-factor-binding site of the ribosome, explaining its elongation arrest activity.

The existence of a signal sequence ‘binding factor’ for protein targeting to the endoplasmic reticulum was first proposed¹ in 1971. This binding factor was subsequently identified in a mammalian system as an 11S ribonucleoprotein (RNP) particle named the SRP². This particle shows three main activities in the process of cotranslational targeting: first, it binds to signal sequences emerging from the translating ribosome; second, it pauses peptide elongation; and third, it promotes protein translocation by docking to the membrane-bound SRP receptor and transferring the ribosome nascent chain complex (RNC) to the protein-conducting channel³.

These activities can be assigned to the two main domains of SRP that are separable by micrococcal nuclease treatment⁴. The first domain, the S domain, binds to signal sequences and promotes translocation⁵. It includes about half of the 7S RNA of SRP (roughly nucleotides 100–250), as well as the essential proteins SRP19, SRP54 and the SRP68–SRP72 (SRP68/72) heterodimer. Although SRP19 is required for SRP assembly⁶, SRP54 is the functionally most significant protein subunit of the S domain: it recognizes the signal sequence⁵ and interacts with the SRP receptor in a GTP-dependent manner⁷. SRP54 comprises an amino-terminal domain (N), a central GTPase domain (G) and a methionine-rich carboxy-terminal domain (M)⁸, which anchors SRP54 to SRP RNA⁹. In addition, together with part of the RNA backbone¹⁰, the M domain carries out the principal function of signal sequence recognition¹¹ near the peptide exit site of the large ribosomal subunit¹².

The second domain of SRP, the Alu domain, mediates the elongation arrest activity¹³. It is supposed to allow efficient targeting by providing a time window in which the nascent chain can be targeted to the translocation site^{14–16}. The Alu domain contains the 5′ and 3′ parts of 7S RNA (including the *Alu*-like sequences), as well as the SRP9–SRP14 (SRP9/14) heterodimer, which is essential for its activity¹⁷.

Little is known about the structural arrangement of the complete SRP^{18,19}, especially when bound to the active ribosome. How can SRP recognize a signal sequence and stop elongation at the same time? We have determined the structure of mammalian SRP bound

to an elongation-arrested 80S ribosome bearing a nascent polypeptide chain containing a signal sequence at 12 Å resolution by using cryo-electron microscopy (cryo-EM) and single-particle reconstruction.

RNC purification and RNC–SRP reconstitution

Because the formation of a stable complex was a prerequisite for our study, we used wheat germ RNCs and canine SRP to reconstitute the targeting complex. This well-characterized combination, which led to the discovery of SRP, shows strong elongation arrest activity¹⁴. Assuming that this activity is a result of equally stable binding of the S and Alu domains to the ribosome, we considered that the wheat germ–canine heterologous complex was the most suitable candidate for structure determination.

We first isolated programmed ribosomes carrying a functional signal sequence (RNCs) from an *in vitro* translation reaction²⁰. The nascent chain represented the first 90 amino acids of the type II membrane protein dipeptidylpeptidase B (DPAP-B) from yeast, which contains a signal anchor sequence, and also a haemagglutinin (HA) and histidine tag. Stalled RNCs were affinity purified and used for reconstitution with excess amounts of purified canine SRP (see Supplementary Information). To ensure specificity—that is, signal-sequence-dependent complex formation—we carried out sucrose density gradient centrifugation under high-salt conditions (500 mM potassium acetate)²¹, which confirmed high-affinity binding of SRP to RNCs with an estimated occupancy of between 50 and 90% (see Supplementary Information).

Cryo-EM map and model of mammalian SRP

Cryo-EM and three-dimensional (3D) reconstruction of the targeting complex shows the typical appearance of an 80S ribosome at 12 Å resolution (7.7 Å according to the 3σ criterion; see Supplementary Information) with two additional densities (Fig. 1). First, a transfer RNA is visible in the P site in the ribosomal intersubunit space. Second, a large elongate mass representing SRP stretches from the peptide exit site of the 60S ribosomal subunit (S domain)

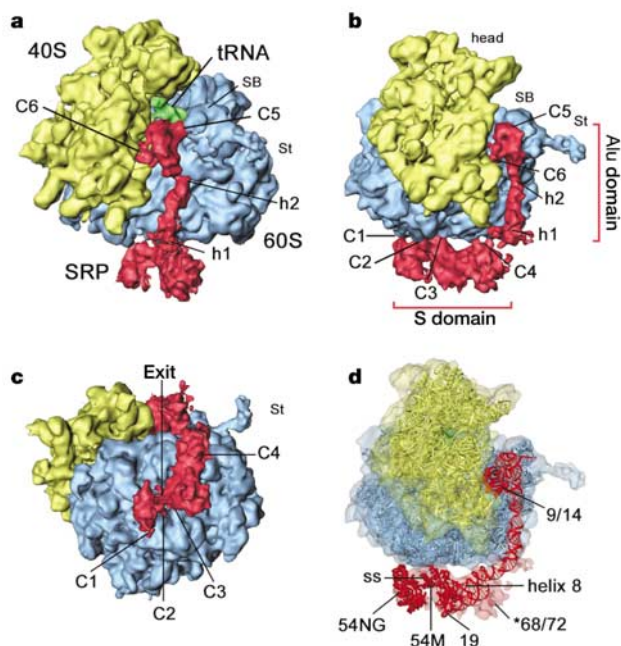


Figure 1 Cryo-EM map of mammalian SRP bound to 80S RNC at 12.0 Å. **a**, RNC-SRP map showing the separated colour-coded densities. The 40S small ribosomal subunit is shown in yellow, 60S large ribosomal subunit in blue, P-site tRNA in green, and SRP in red. C1–C6 indicate the assigned positions of RNC-SRP connections (see Table 1); h1 and h2 are hinges of the 7S RNA backbone of SRP; St, stalk; SB, stalk base. **b**, As **a**, but rotated by 70° to the right. **c**, As **a**, but rotated upwards by 90°. **d**, Same orientation as **b**, but with molecular models of SRP and 80S RNC shown in transparent densities.

into the intersubunit space (Alu domain), forming a total of six connections (C1–C6) with the ribosome (Fig. 1).

The tRNA density reflects the presence of the nascent peptidyl-tRNA containing the signal sequence, which is stalled at the 3' end of the truncated messenger RNA and stabilized by cycloheximide²⁰. In agreement with the occupancy estimated by SDS polyacrylamide gel electrophoresis (SDS-PAGE), sorting the data set according to the presence of SRP density resulted in a subset of about 70% of the particles, which were used in the final reconstruction.

To facilitate interpretation on a molecular level, we attempted to dock molecular models into the electron densities (Fig. 2). For the ribosome, the high similarity between wheat germ RNC and yeast

RNC²⁰ (Figs 3 and 4) allowed the use of a previously generated molecular model²². Therefore, we use the yeast nomenclature for the molecular description of ribosomal components (with family names given in parentheses). For SRP we used X-ray structures of SRP fragments, which were docked into the density as rigid bodies (see Methods).

SRP has a bent conformation, with one of two hinges apparently facilitating a major kink (hinge 1) separating the S and Alu domains, in agreement with a three-domain structure of length 260–280 Å that has been proposed for SRP in solution¹⁸. Hinge 1 separates the 160-Å S domain of SRP near the peptide exit site from a linker connecting the Alu domain in a region close to the subunit interface (spanning a total length of 120 Å). The RNA at hinge 1 represents a large loop around nucleotides 100 and 250, and forms an angle of almost 90° (Figs 1b and 2). This site coincides precisely with the cutting site for micrococcal nuclease⁴, which separates SRP into its main domains. Hinge 2 is located in a region corresponding to a small loop formed by nucleotides 70 and 275 of 7S RNA. This hinge facilitates a bend (~30°) that leads to an orientation of the Alu 5' RNP that is perfect for its entry into the intersubunit space (Figs 1a and 2f).

Into the density identified as the S domain, we docked the structure of a large fragment of the mammalian S domain containing 7S RNA helices 6–8 and part of helix 5, as well as the SRP19 protein and the SRP54 M domain²³. The original SRP54 M domain was replaced with another model²⁴, which differs only in the position of helix 1 and the finger loop. As a signal sequence, we positioned an α-helical peptide fragment into the corresponding density near the exit site. As a result, it can contact the hydrophobic groove of the SRP54 M domain and the phosphate backbone of the SRP helix 8 RNA¹⁰. Next, we docked the conserved structure of a prokaryotic SRP54 NG domain²⁵. Notably, the NG domain is positioned such that a gap of about 20 Å is separating it from helix 8 of 7S RNA, which is only connected by the M domain. In this position, part of the NG domain is too close to the finger loop of the M domain, indicating that there must be a more compact conformation of the loop for signal sequence binding. In support of our model, the crystal structure of an archaeal SRP54–RNA complex shows a similar overall arrangement²⁶.

We interpreted extra density in the S domain as the SRP68/72 dimer of a hitherto unknown structure. It is located mainly at the junction of helices 5–8; however, additional density is present at the hinge 1 region of helix 5 (Fig. 2c). This is in accordance with footprinting experiments showing that all of these regions of 7S RNA (including nucleotides 100 and 250) are protected²⁷. The fragmented mass may be an indication of a tertiary structure

Table 1 Contacts between mammalian SRP and the 80S ribosome

Contact	Type*	SRP	RNA/protein position	80S ribosome†	RNA/protein position‡
C1	p–p	S domain	60–75	60S subunit	130–135
		54 NG		rpL25 (L23p)	
C2	p/R–p	54 NG	15–26	rpL35 (L29p)	16–27
		54 M/H8		Signal sequence	
C3	p–R	54 M	388–399	H59/ES24	1,627–1,634§
		54 M		H24	
C4	R–R/p	H5	218–228, 121–127	H99, H100/101, rpL16	2,907–2,910, 2,849–2,851
		68/72		H98/ES39, rpL16 (L13p)	
C5	R–p	Alu domain	135–136	60S subunit	65–69
		L1.2		rpL12 (L11p)	
		L2		H43	
		H2		H95/SRL	
C6	p–R	14	74–89	40S subunit	55–58, 356–359; 368
		9		h5, h15	
		9		h5, h15	
		9		h14	

*R and p correspond to RNA and protein, respectively.

†Yeast nomenclature is used with family name given in parentheses. ES, expansion segment.

‡Positions correspond to a model based on the yeast 80S ribosome²².

§Positions correspond to yeast 25S RNA secondary structure (<http://www.mla.icmb.utexas.edu/>).

containing thin, extended 'tentacle' regions, as observed for some ribosomal proteins such as L19 or L22 (ref. 28). In the described position, SRP68/72 can function as a brace between the core of the S domain and the dynamic hinge 1, thereby functionally connecting the Alu and S domains.

The X-ray structure of the mammalian SRP9/14 dimer bound to the 5' part of the Alu RNA²⁹ fits perfectly into the density in the intersubunit space (Fig. 2). This model matches a previously suggested conformation²⁹, in which the 5' RNP, comprising the SRP9/14 dimer and the first 48 nucleotides of the RNA, folds back onto the 3' RNA stem of the Alu domain. Thus, this back-folding seems to be a necessary assembly step of the Alu domain. In a final step, we used three fragments from a model provided by the SRP database³⁰ as a ruler to corroborate that the missing part of 7S RNA

can span the distance between the docked Alu and S domain fragments.

Taken together, density consistent with size and structure could be assigned to all known components of the mammalian SRP, thereby leading to the first molecular model of SRP in the functional context of a ribosomal targeting complex.

An overview of the complete model in the context of the ribosome is shown in Figs 1d and 2e, f. The docked fragments easily span the distance between the peptide exit site and the binding site for elongation factors. The three connections of the S domain with the ribosome, found in the immediate vicinity of the peptide exit site, are contributed exclusively by the SRP54 protein. The Alu domain bridges, in a tight fit, the ~65 Å between the large and small ribosomal subunits (Fig. 2f). An empty space in the intersubunit

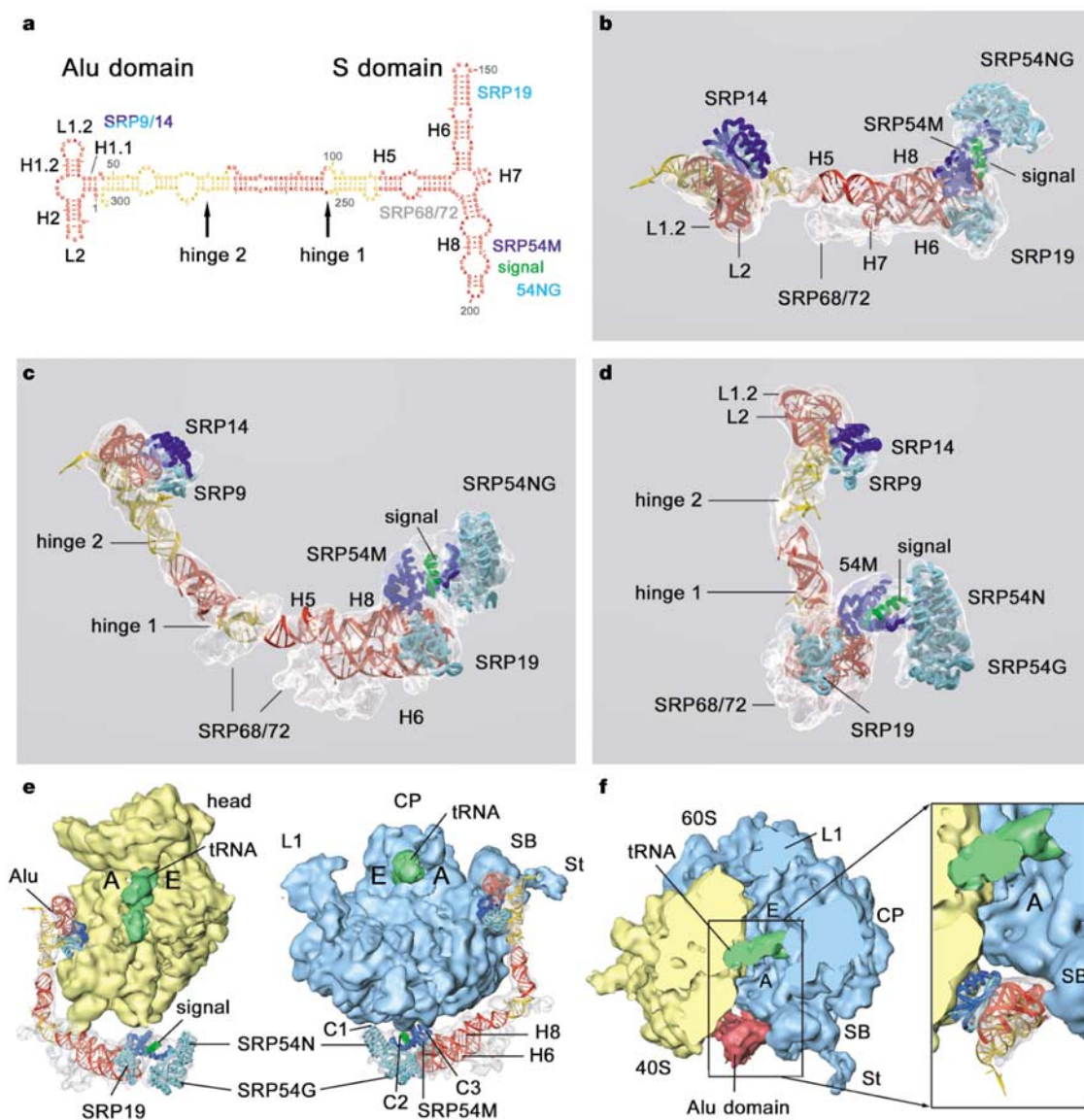


Figure 2 Molecular model of SRP. **a**, Secondary structure of SRP RNA with protein-binding sites and hinges indicated. H1–H8 denote SRP RNA helices, numbered according to ref. 29 for the Alu domain. SRP proteins are shown in cyan, blue and grey, 7S RNA in red and yellow, and the signal sequence in green. **b**, Molecular model of SRP with transparent density and colour coding as in **a**. Top view, SRP as seen from the ribosome. **c**, As **b**, but rotated upwards. **d**, As **c**, but rotated left. **e**, SRP with isolated 40S and 60S

ribosomal subunits exposing the P-site tRNA (green), shown from the 60S (left) and the 40S (right) side. A and E, positions of the A and E sites; Alu, Alu domain; head, head of 40S subunit; L1, L1 protuberance; CP, central protuberance. Other labels are the same as in **a**. **f**, Cut top view showing the Alu domain in the intersubunit space (left), labelled as in **e**, and magnified view showing a molecular model of the Alu domain coloured as in **a**.

cavity, corresponding to the unoccupied A site, indicates that the Alu domain will not interfere with a tRNA bound in this position (Fig. 2e, f).

Environment and function of the S domain

The first connection between the S domain and the large ribosomal subunit is formed by the tip of the SRP54 N domain (Fig. 3a). Its two loops, which connect the four-helix bundle, come into close proximity with rpL25 and rpL35 (rpL25/35, corresponding to L23a/L35 in wheat germ and L23p/L29p in *Escherichia coli*). This is in agreement with previous crosslinking experiments identifying the same proteins as the main proteinaceous ribosomal constituents in immediate vicinity to SRP54 (ref. 12). In similar experiments, the same region of the bacterial SRP54 N domain has been found in a position adjacent to L23p³¹, suggesting that this interaction is evolutionarily conserved. In addition to SRP, signal sequences³² and the chaperone trigger factor^{33,34} bind to L23p, and the protein-conducting channel (PCC) of the endoplasmic reticulum (the Sec61 complex) binds to rpL25/35 (ref. 20 and Fig. 3g). Thus, the rpL25/35 (L23p/29p) proteins constitute a promiscuous binding site of the ribosome that facilitates interaction with several factors involved in different aspects of cotranslational processing.

The second connection is formed by the N-terminal part of the SRP54 M domain contacting helix 59 (expansion segment 24) of the 25S ribosomal RNA (Fig. 3b, c). As in the first connection involving rpL25/35, this contact site of SRP is shared with the Sec61 complex²⁰. The signal sequence is also closest to this connection when bound to SRP in the suggested position.

The third connection engages the C-terminal region of the M domain, in particular the part that binds 7S RNA, which, again like

the Sec61 complex, interacts with helix 24 of the 25S ribosomal RNA (Fig. 3b, c). However, in contrast to the Sec61 complex, which contacts the stem of this helix, the SRP-binding site is shifted towards the tip of helix 24. The exclusive involvement of ribosomal RNA in connections 2 and 3 is in agreement with the observation that rpL25/35 (L23p/29p) are the only ribosomal proteins that are crosslinked to SRP54.

Forming the described contacts with the ribosome, SRP54 assumes an open conformation (Fig. 3c, e). How does this relate to the signal-sequence-dependent interaction of SRP with the ribosome? We speculate that in a first step of the ribosome–SRP interaction, which is independent of the signal sequence and involves at least connection 1 (ref. 12), SRP54 switches to a conformation that is competent for signal sequence scanning (sampling mode). In a second step, interaction with a functional signal sequence may induce and stabilize the observed open conformation of SRP54, enabling SRP to bind with high affinity to the ribosome (targeting mode). Notably, the N and the M domains of SRP54 are positioned on the ribosome, thereby determining the orientation of the G domain in between them. As a result, signal sequence binding to the M domain could regulate the nucleotide affinity of the G domain by influencing its position relative to the N domain, which is thought to control nucleotide affinity and GTPase activity²⁵.

The fourth connection is the only one involving 7S RNA of the S domain (Fig. 3c, d). It seems to be split into two entities, one of which (connection 4a) is very close to helix 5 of SRP; however, it is likely that SRP68/72 is also involved in this connection (4b). The second connection (4b) is located right next to a ribosomal density that has been identified as expansion segment 39 (ref. 22), formed

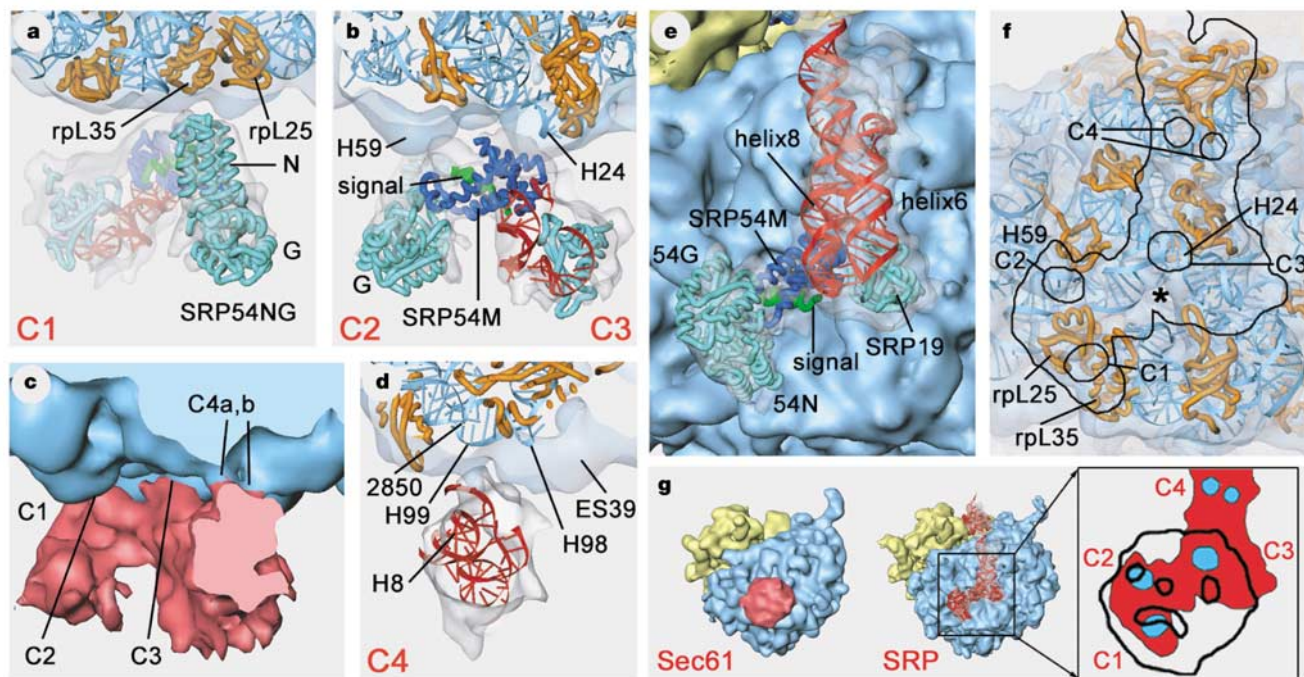


Figure 3 Interaction of the SRP S domain with the 80S ribosome. **a**, Density with molecular models inserted and cut to show connection 1 (C1) between the SRP54 NG domain and rpL25/35. SRP is coloured as in Fig. 2, with 60S ribosomal density and 25S RNA shown in light blue, and 60S proteins in orange. **b**, Connections 2 and 3 (C2, C3) between the SRP54 M domain and ribosomal helices H59 and H24, respectively. **c**, Density in the same orientation as **b** and **d**, cut to show connections C2–C4. **d**, Connection 4 between helix 5 of SRP (and SRP68/72) and the 60S subunit. 2850, position of nucleotide 2,850 in the 25S ribosomal RNA model; H99 and H98, 25S RNA

helices; ES39, expansion segment 39. **e**, Model of the S domain covering the peptide exit site of the 60S subunit. **f**, As **e**, with transparent ribosomal density and contour of SRP density showing locations of the connections. Asterisk denotes the peptide tunnel exit. **g**, Comparison of the 80S RNC–Sec61 complex from yeast and the 80S RNC–SRP complex in the same orientation. Magnified area shows the contours of SRP (red) and Sec61 (black line), and their partially overlapping contact sites: SRP (blue), Sec61 (black line).

by an extension of helix 98 of 25S RNA. In addition, rpL16 (L13p) projects a loop into the vicinity of connection 4. The closest ribosomal structures found for connection 4a are helix 99 and a small loop at the junction between helices 100 and 101 of 25S RNA. Notably, the corresponding region in the *E. coli* ribosome, around nucleotide 2,828 of 23S RNA (nucleotide 2,850 in our model), has been found in the vicinity of bacterial SRP 4.5S RNA by cross-linking³⁵, again suggesting that there is a conserved mode of interaction.

Thus, considering the previously mentioned localization of SRP54, the core of the eukaryotic S domain, including SRP54 and helices 8 and 5, seems to be positioned on the ribosome in an overall orientation similar to that of its counterpart in prokaryotes.

When comparing the binding sites of SRP and PCC^{20,36}, it is evident that both cannot bind to the ribosome at the same time (ref. 37 and Fig. 3g). Therefore, docking of the ribosome to the PCC first requires a rearrangement of the whole S domain relative to the ribosome, which is triggered by interaction of the SRP–RNC complex with the SRP receptor¹². This rearrangement may precede a state in which the binding sites for PCC are accessible and the transfer of the signal sequence can take place.

Environment and function of the Alu domain

The two connections between the Alu domain and the ribosome are contributed exclusively by the 5' RNP comprising the first 48 nucleotides and the SRP9/14 heterodimer (Fig. 4). In connection 5, the 5' RNA of the SRP is interacting with both RNA and protein of the large ribosomal subunit (Fig. 4b). Loops L1.2 and L2, as well as the short helix 2 of 7S RNA, contact the large ribosomal subunit through the so-called 'stalk base', and probably through the universally conserved α -sarcin–ricin loop. The participating com-

ponents of the stalk base are the N-terminal part of rpL12 (L11p) and the tip of helix 43 of 25S ribosomal RNA.

In connection 6, the only contact of SRP with the small ribosomal subunit is established through the SRP9/14 dimer and ribosomal 18S RNA (Fig. 4a). The SRP14 surface mainly participates in contacts with helices 5 and 15 of 18S RNA. SRP9 is in contact with the same helices and, in addition, is close to helix 14. The functionally essential C terminus³⁸ and a large loop between the β 2 and β 3 strands of SRP14 are not resolved in the X-ray structure²⁹; therefore, at the given resolution we cannot draw any conclusions regarding their participation in ribosomal contacts.

The ribosomal components bound by the Alu domain are well conserved in all ribosomes and comprise the elongation-factor-binding site³⁹. It is intriguing that all of the contact sites used by the SRP Alu domain are also used by eEF2 (refs 40, 41), enabling us to interpret the Alu–ribosome interaction as elongation factor mimicry (Fig. 4c). A tRNA-like interaction between loop L2 of the SRP 5' RNP and helix 43 of the stalk base is reminiscent of the interaction between the tRNA T-loop and the same ribosomal helix in the context of the EF–Tu–tRNA–GTP ternary complex bound to the ribosome⁴².

Binding of the Alu domain in this position directly competes with elongation factors entering their binding site (Fig. 4c). Therefore, sufficiently high affinity of the Alu domain for this site explains its elongation arrest activity. Variations in Alu-domain affinity could explain the different efficiencies observed in different systems^{16,43}. In our structure, the Alu domain is bound to a ribosome in the post-translocational conformation ('post' state), as defined by the presence of the peptidyl–tRNA in the P site and an unoccupied A site. Thus, it is possible that the post state of the ribosome is the preferred

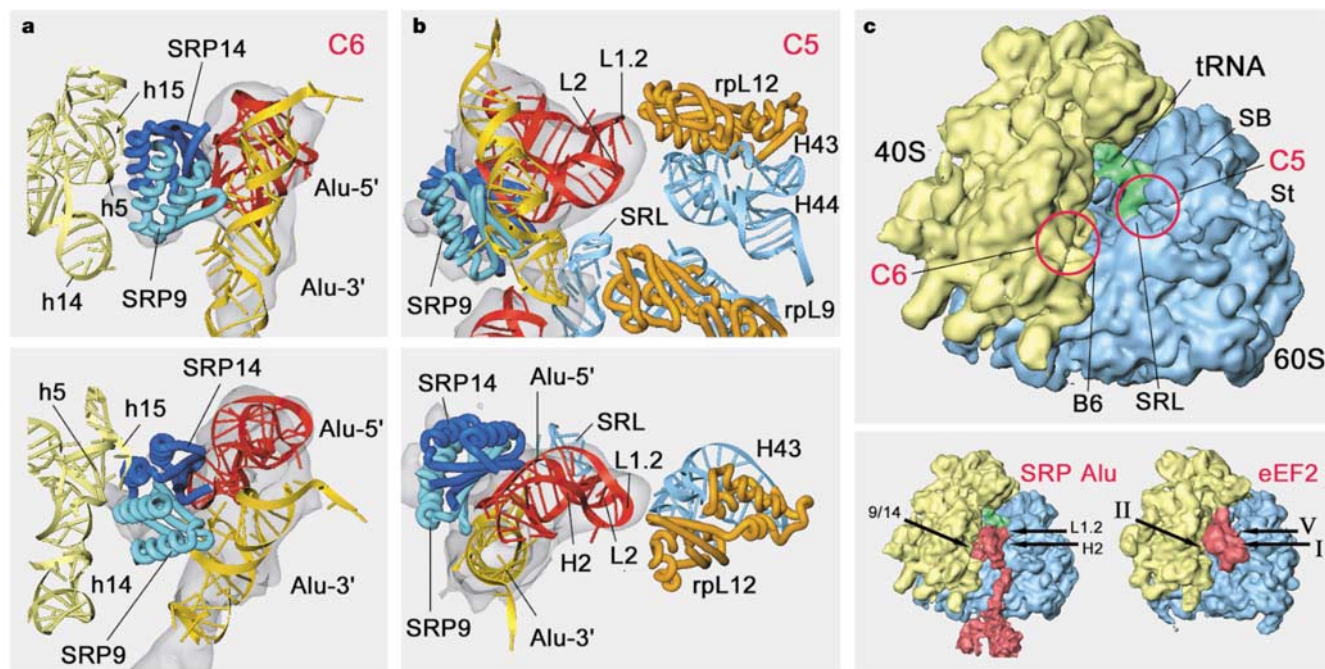


Figure 4 Interaction of the SRP Alu domain with the 80S ribosome. **a**, Top, SRP density with models showing connection 6 (C6) between SRP9/14 and ribosomal 18S RNA (helices 5, 15 and 14) in an orientation similar to Fig. 1b. Colour coding is the same as in Figs 2 and 3, with 18S ribosomal RNA shown in pale yellow. Bottom, as top view, but tilted towards the viewer. **b**, Top, connection 5 (C5) between SRP Alu RNA (loop 1.2 and helix 2) and the GAC of the 60S ribosomal subunit (rpL12, helix 43 and the α -sarcin–ricin loop (SRL)). Bottom, as top view, but tilted towards the viewer. **c**, Top, 80S RNC density without SRP, showing conserved contact sites of the Alu domain. B6, bridge 6. Bottom, comparison of the 80S RNC–SRP complex and the 80S ribosome–eEF2 complex⁴¹ in the same orientation (SRP and eEF2 are shown in red). Subunits of the Alu domain and domains of eEF2 involved in similar contact sites are labelled.

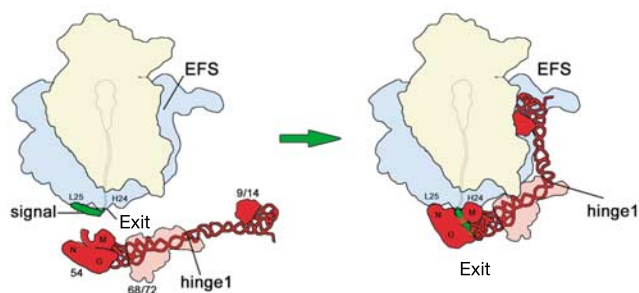


Figure 5 Signal-sequence-dependent SRP-ribosome interaction. On signal sequence binding by SRP54, a kinked conformation of SRP involving possibly SRP68/72 and a rotation around hinge 1 is stabilized. As a result, SRP interacts with the ribosome, stretching from the peptide exit (S domain) to the elongation-factor-binding site in the intersubunit space (Alu domain), where it causes elongation arrest by competition with elongation factors. Colour coding is the same as in Fig. 1, with the signal sequence (signal) shown in green. Exit, peptide tunnel exit; EFS, elongation-factor-binding site.

conformation for SRP binding during the elongation cycle⁴⁴.

How can elongation arrest by the Alu domain be controlled by the event of signal sequence binding, which occurs more than 250 Å away? Signal sequence binding has to induce or at least to stabilize bending⁴⁵ of the particle at hinge 1 to promote Alu-domain binding (Fig. 5). One possibility is that high-affinity binding of the S domain simply tethers the Alu domain in an appropriate position on the ribosome, and thereby favours the bound-state conformation of SRP. However, we prefer another model in which the recognition of a signal peptide produces specific positioning of the S domain on the ribosome; this may lead to conformational changes in SRP68/72 that result in hinge 1 stabilizing a 90° angle and the Alu domain closing into the elongation-factor-binding site (Fig. 5). In agreement with this idea is the brace-like localization of SRP68/72 covering the hinge 1 region, its participation in connection 4, and the finding that SRP reconstituted without SRP68/72 lacks elongation arrest activity¹⁷.

Conclusions

The structure of the eukaryotic SRP-RNC targeting complex shows that SRP accomplishes its task of signal sequence recognition and elongation pausing by spanning from the peptide exit site to the elongation-factor-binding site of the ribosome in a kinked conformation. The ribosomal contacts involve 7S RNA and all SRP proteins, with the exception of SRP19. SRP54, as the main protagonist of the S domain, establishes an open conformation directly across the peptide exit site and SRP68/72 may communicate between the S and the Alu domains. The position of the Alu domain in the elongation-factor-binding site explains its elongation arrest activity by direct competition. The structure suggests that the conserved core of the S domain of SRP is likely to follow the same mode of ribosome interaction and functioning in all organisms. □

Methods

Purification of RNCs

To generate purified RNCs, we used a wheat germ *in vitro* translation system (Ambion) programmed with truncated mRNA encoding the 90 N-terminal amino acids of DPAP-B from *Saccharomyces cerevisiae*. Purification was done by a modified protocol²⁰. A DNA fragment with N-terminal histidine and HA tags was generated by polymerase chain reaction from yeast genomic DNA using forward (5'-taatcagctactataggacacaa caaaacaaataaaacacacaaatgtctcatcatcatcatcattaccatagatgtccagattacgtgaaggtggcgaag aagaattgt-3') and reverse (5'-ttgcagctctgatatttggatg-3') primers. We then synthesized capped mRNA by a Message Machine kit (Ambion). For translation, six 200-μl reactions were incubated for 45 min at 27 °C and terminated with 2 μl of 10 mg ml⁻¹ cycloheximide.

Reactions were spun through a high-salt sucrose cushion (50 mM Tris-HCl (pH 7.0), 500 mM potassium acetate, 25 mM magnesium acetate, 2 mM dithiothreitol (DTT), 1 M sucrose and 10 μg ml⁻¹ cycloheximide) at 355,000g for 45 min. Pellets were resuspended in ice-cold 250 buffer (50 mM Tris-HCl (pH 7.0), 250 mM potassium acetate, 25 mM

magnesium acetate, 0.1% (w/v) Nikkol, 5 mM β-mercaptoethanol, 10 μg ml⁻¹ cycloheximide and 250 mM sucrose) and transferred to 1.5 ml of Talon metal-affinity resin (Clontech). The resin was washed with 8 ml of 250 buffer, and 2 ml of 500 buffer (250 buffer containing 500 mM potassium acetate). RNCs were eluted with 100 mM imidazol (pH 7.1) in 250 buffer and spun through 500 μl of a high-salt sucrose cushion. The resulting pellet was slowly resuspended in G buffer (20 mM Tris-HCl (pH 7.0), 50 mM potassium acetate, 10 mM magnesium acetate, 1 mM DTT, 125 mM sucrose, 100 μg ml⁻¹ cycloheximide, 0.05% (w/v) Nikkol and 0.03% (w/v) of an EDTA-free complete protease inhibitor pill (Boehringer) and 0.2 U μl⁻¹ RNasin (Ambion)), flash frozen and stored at -80 °C. From 1.2 ml of translation reaction, RNCs with an absorbance of 0.7 at 260 nm (~15 pmol) were isolated.

Reconstitution of the SRP-RNC complex

RNC-SRP complexes were reconstituted by incubating 1.5 pmol of mammalian SRP (isolated as described⁴⁶ and further purified by sucrose density gradient centrifugation⁴⁷) and 0.5 pmol of RNCs for 15 min at room temperature in 25 mM HEPES (pH 7.5), 150 mM potassium acetate, 5 mM DTT, 5 mM magnesium acetate, 100 mM sucrose, 0.02% Nikkol, 100 μg ml⁻¹ cycloheximide and 0.06% of an EDTA-free complete protease inhibitor pill. The reaction was spun through a 10–40% high-salt sucrose cushion (500 mM potassium acetate) and analysed by SDS-PAGE. For cryo-EM, 1 pmol of RNCs was incubated with 3 pmol of SRP in a volume of 28 μl under the described conditions, except that the potassium acetate concentration was 180 mM.

Electron microscopy, image processing and models

Samples were applied to carbon-coated holey grids as described⁴⁸. Micrographs were recorded under low-dose conditions on a Tecnai F30 field emission gun electron microscope at 300 kV and a Tecnai F20 instrument at 160 kV in a defocus range of 1.0 μm to 4.5 μm, and scanned on a Heidelberg drum scanner, resulting in a pixel size of 3.26 Å on the object scale. The data were analysed by the SPIDER software package. After automated particle picking followed by visual inspection, 35,488 particles were selected for density reconstruction. We sorted the data set into subsets (±SRP) according to a procedure developed by C.M.T.S., P. Penczek and J.F. (unpublished). Removal of particles lacking the SRP resulted in two subsets of 10,091 (–SRP) and 25,397 (+SRP) particles, which were used for the final CTF-corrected reconstruction at a resolution of 12.0 Å (7.7 Å), based on the Fourier shell correlation with a cutoff value of 0.5 (3σ). Densities for the 40S subunit, the 60S subunit, the P-site tRNA and the SRP were isolated by using binary masks. Amplitude correction was done by Fourier filtering using B factors. A lower contour level of the SRP density for surface representation was applied. This indicates that the SRP density is under-represented because of incomplete removal of SRP-free ribosomal particles from the final particle subset (the same contour level is shown in Supplementary Fig. 1).

Docking of X-ray structures and molecular models of SRP was done by the programs SPIDER and O⁴⁹. First, a fragment of the mammalian S domain containing 7S RNA helices 6–8, part of helix 5, SRP19 and the SRP54 M domain (Protein Data Bank (PDB) accession code 1MFQ; ref. 23) was docked. The M domain was replaced by a different model²⁴ using the RNA-binding moiety for alignment. The structure of a prokaryotic SRP54 NG domain (PDB 1JPI; ref. 25) was docked into density present near the M domain. An α-helical peptide fragment was docked as the signal sequence. The X-ray structure of the mammalian Alu 5' RNP (PDB 1E8O; ref. 29) was docked and, for the missing part of the 7S RNA, three fragments were used from a model provided by the SRP database³⁰. The high degree of similarity between the wheat germ and the yeast RNC allowed us to use a molecular model of the yeast ribosome (PDB 1K5X, 1K5Y and 1K5Z; ref. 22). The figures were prepared by using Iris Explorer (NAG), Ribbons⁵⁰ and POV-Ray.

Received 21 November 2003; accepted 16 January 2004; doi:10.1038/nature02342.

1. Blobel, G. & Sabatini, D. in *Biomembranes* (ed. Manson, L. A.) 193–195 (Plenum, New York, 1971).
2. Walter, P., Ibrahim, I. & Blobel, G. Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to *in-vitro*-assembled polysomes synthesizing secretory protein. *J. Cell Biol.* **91**, 545–550 (1981).
3. Koch, H. G., Moser, M. & Muller, M. Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. *Rev. Physiol. Biochem. Pharmacol.* **146**, 55–94 (2003).
4. Gundelfinger, E. D., Krause, E., Melli, M. & Dobberstein, B. The organization of the 7S RNA in the signal recognition particle. *Nucleic Acids Res.* **11**, 7363–7374 (1983).
5. Siegel, V. & Walter, P. Each of the activities of signal recognition particle (SRP) is contained within a distinct domain: analysis of biochemical mutants of SRP. *Cell* **52**, 39–49 (1988).
6. Walter, P. & Blobel, G. Disassembly and reconstitution of signal recognition particle. *Cell* **34**, 525–533 (1983).
7. Connolly, T. & Gilmore, R. The signal recognition particle receptor mediates the GTP-dependent displacement of SRP from the signal sequence of the nascent polypeptide. *Cell* **57**, 599–610 (1989).
8. Bernstein, H. D. et al. Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature* **340**, 482–486 (1989).
9. Romisch, K., Webb, J., Lingelbach, K., Gausepohl, H. & Dobberstein, B. The 54-kD protein of signal recognition particle contains a methionine-rich RNA binding domain. *J. Cell Biol.* **111**, 1793–1802 (1990).
10. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B. & Doudna, J. A. Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* **287**, 1232–1239 (2000).
11. Zopf, D., Bernstein, H. D., Johnson, A. E. & Walter, P. The methionine-rich domain of the 54 kD protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence. *EMBO J.* **9**, 4511–4517 (1990).
12. Pool, M. R., Stumm, J., Fulga, T. A., Sinning, I. & Dobberstein, B. Distinct modes of signal recognition particle interaction with the ribosome. *Science* **297**, 1345–1348 (2002).
13. Siegel, V. & Walter, P. Removal of the Alu structural domain from signal recognition particle leaves its

- p>protein translocation activity intact.
- Nature*
- 320**
- , 81–84 (1986).
14. Walter, P. & Blobel, G. Translocation of proteins across the endoplasmic reticulum. III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J. Cell Biol.* **91**, 557–561 (1981).
 15. Wolin, S. L. & Walter, P. Signal recognition particle mediates a transient elongation arrest of preprolactin in reticulocyte lysate. *J. Cell Biol.* **109**, 2617–2622 (1989).
 16. Mason, N., Ciufu, L. F. & Brown, J. D. Elongation arrest is a physiologically important function of signal recognition particle. *EMBO J.* **19**, 4164–4174 (2000).
 17. Siegel, V. & Walter, P. Elongation arrest is not a prerequisite for secretory protein translocation across the microsomal membrane. *J. Cell Biol.* **100**, 1913–1921 (1985).
 18. Andrews, D. W., Walter, P. & Ottensmeyer, F. P. Evidence for an extended 7SL RNA structure in the signal recognition particle. *EMBO J.* **6**, 3471–3477 (1987).
 19. Nagai, K. *et al.* Structure, function and evolution of the signal recognition particle. *EMBO J.* **22**, 3479–3485 (2003).
 20. Beckmann, R. *et al.* Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* **107**, 361–372 (2001).
 21. Walter, P. & Blobel, G. Subcellular distribution of signal recognition particle and 7SL-RNA determined with polypeptide-specific antibodies and complementary DNA probe. *J. Cell Biol.* **97**, 1693–1699 (1983).
 22. Spahn, C. M. *et al.* Structure of the 80S ribosome from *Saccharomyces cerevisiae* tRNA–ribosome and subunit–subunit interactions. *Cell* **107**, 373–386 (2001).
 23. Kuglstatter, A., Oubridge, C. & Nagai, K. Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nature Struct. Biol.* **9**, 740–744 (2002).
 24. Huang, Q., Abdulrahman, S., Yin, J. & Zwieb, C. Systematic site-directed mutagenesis of human protein SRP54: interactions with signal recognition particle RNA and modes of signal peptide recognition. *Biochemistry* **41**, 11362–11371 (2002).
 25. Padmanabhan, S. & Freymann, D. M. The conformation of bound GMPPNP suggests a mechanism for gating the active site of the SRP GTPase. *Struct. Fold. Des.* **9**, 859–867 (2001).
 26. Rosendal, K. R., Wild, K., Montoya, G. & Sinning, I. Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc. Natl Acad. Sci. USA* **100**, 14701–14706 (2003).
 27. Siegel, V. & Walter, P. Binding sites of the 19-kDa and 68/72-kDa signal recognition particle (SRP) proteins on SRP RNA as determined in protein–RNA ‘footprinting’. *Proc. Natl Acad. Sci. USA* **85**, 1801–1805 (1988).
 28. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
 29. Weichenrieder, O., Wild, K., Strub, K. & Cusack, S. Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature* **408**, 167–173 (2000).
 30. Rosenblad, M. A., Gorodkin, J., Knudsen, B., Zwieb, C. & Samuelsson, T. SRPDB: signal recognition particle database. *Nucleic Acids Res.* **31**, 363–364 (2003).
 31. Gu, S. Q., Peske, F., Wieden, H. J., Rodnina, M. V. & Wintermeyer, W. The signal recognition particle binds to protein L23 at the peptide exit of the *Escherichia coli* ribosome. *RNA* **9**, 566–573 (2003).
 32. Eisner, G., Koch, H. G., Beck, K., Brunner, J. & Muller, M. Ligand crowding at a nascent signal sequence. *J. Cell Biol.* **163**, 35–44 (2003).
 33. Kramer, G. *et al.* L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419**, 171–174 (2002).
 34. Ullers, R. S. *et al.* Interplay of signal recognition particle and trigger factor at L23 near the nascent chain exit site on the *Escherichia coli* ribosome. *J. Cell Biol.* **161**, 679–684 (2003).
 35. Rinke-Appel, J. *et al.* Crosslinking of 4.5S RNA to the *Escherichia coli* ribosome in the presence or absence of the protein Ffh. *RNA* **8**, 612–625 (2002).
 36. Morgan, D. G., Menetret, J. F., Neuhof, A., Rapoport, T. A. & Akey, C. W. Structure of the mammalian ribosome–channel complex at 17 Å resolution. *J. Mol. Biol.* **324**, 871–886 (2002).
 37. Moller, I. *et al.* A general mechanism for regulation of access to the translocon: competition for a membrane attachment site on ribosomes. *Proc. Natl Acad. Sci. USA* **95**, 13425–13430 (1998).
 38. Thomas, Y., Bui, N. & Strub, K. A truncation in the 14 kDa protein of the signal recognition particle leads to tertiary structure changes in the RNA and abolishes the elongation arrest activity of the particle. *Nucleic Acids Res.* **25**, 1920–1929 (1997).
 39. Wilson, D. N. *et al.* Protein synthesis at atomic resolution: mechanistics of translation in the light of highly resolved structures for the ribosome. *Curr. Protein Peptide Sci.* **3**, 1–53 (2002).
 40. Gomez-Lorenzo, M. G. *et al.* Three-dimensional cryo-electron microscopy localization of EF2 in the *Saccharomyces cerevisiae* 80S ribosome at 17.5 Å resolution. *EMBO J.* **19**, 2710–2718 (2000).
 41. Spahn, C. *et al.* Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.* (in the press).
 42. Valle, M. *et al.* Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nature Struct. Biol.* **10**, 899–906 (2003).
 43. Wolin, S. L. & Walter, P. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* **7**, 3559–3569 (1988).
 44. Ogg, S. C. & Walter, P. SRP samples nascent chains for the presence of signal sequences by interacting with ribosomes at a discrete step during translation elongation. *Cell* **81**, 1075–1084 (1995).
 45. Andreazzoli, M. & Gerbi, S. A. Changes in 7SL RNA conformation during the signal recognition particle cycle. *EMBO J.* **10**, 767–777 (1991).
 46. Martoglio, B., Hauser, S. & Dobberstein, B. in *Cell Biology: A Laboratory Handbook* (ed. Celis, J. C.) 265–273 (Academic, San Diego, 1997).
 47. Walter, P. & Blobel, G. Signal recognition particle: a ribonucleoprotein required for cotranslational translocation of proteins, isolation and properties. *Methods Enzymol.* **96**, 682–691 (1983).
 48. Wagenknecht, T., Grassucci, R. & Frank, J. Electron microscopy and computer image averaging of ice-embedded large ribosomal subunits from *Escherichia coli*. *J. Mol. Biol.* **199**, 137–147 (1988).
 49. Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 50. Carson, M. Ribbons 2.0. *Appl. Crystallogr.* **24**, 103–106 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Böttcher for help with the F20 cryo-microscope; U. Bach for technical assistance; and B. Dobberstein for support. This work was supported by a grant of the VolkswagenStiftung (to R.B.), by grants from NIH, NSF and HHMI (to J.F.) and by the European Union and the Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin in the context of the Ultra-Structure Network.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.B. (roland.beckmann@charite.de). Coordinates of the atomic model of SRP have been deposited in the PDB under accession number 1RY1. The cryo-EM map has been deposited in the 3D-EM database (EMBL–European Bioinformatics Institute, Cambridge, UK) under accession number EMD-1063.

A large neutral fraction of cosmic hydrogen a billion years after the Big Bang

J. Stuart B. Wyithe¹ & Abraham Loeb²

¹School of Physics, The University of Melbourne, Parkville, Victoria 3010, Australia

²Astronomy Department, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA

The fraction of ionized hydrogen left over from the Big Bang provides evidence for the time of formation of the first stars and quasar black holes in the early Universe; such objects provide the high-energy photons necessary to ionize hydrogen. Spectra of the two most distant known quasars¹ show nearly complete absorption of photons with wavelengths shorter than the Lyman α transition of neutral hydrogen, indicating that hydrogen in the intergalactic medium (IGM) had not been completely ionized at a redshift of $z \approx 6.3$, about one billion years after the Big Bang. Here we show that the IGM surrounding these quasars had a neutral hydrogen fraction of tens of per cent before the quasar activity started, much higher than the previous lower limits^{1,2} of ~ 0.1 per cent. Our results, when combined with the recent inference of a large cumulative optical depth to electron scattering after cosmological recombination³ therefore suggest the presence of a second peak in the mean ionization history of the Universe.

The detection of a Ly α -emitting galaxy at $z \approx 6.56$ led to the claim that the IGM must already be highly ionized at that redshift⁴. However, it was later shown⁵ that the presence of a small ionized region around the galaxy and broadening of the Ly α line would allow detection of the line feature in a fully neutral IGM. Luminous quasars offer an alternative probe of the IGM and have higher luminosities, allowing acquisition of higher-quality spectra. The spectra^{6,7} of SDSS J1030+0524 at $z = 6.28$ and SDSS J1148+5251 at $z = 6.41$ show transmitted flux down to wavelengths corresponding to the Ly α resonance of hydrogen at $z \approx 6.20$ and $z \approx 6.32$ respectively, even though no flux is detected at somewhat greater distances from the quasars⁷. These redshift displacements correspond to velocity shifts of order $\sim 3,300 \text{ km s}^{-1}$, and imply ionized regions surrounding the quasars with observed physical radii of $R_{\text{obs}} = 4.5 \text{ Mpc}$ and 4.7 Mpc , respectively. The inference is robust because high-ionization broad-lines typically have velocity offsets from the true quasar redshift⁸ of only $\lesssim 1,000 \text{ km s}^{-1}$.

We model the evolution of a fully ionized region (the Strömgren sphere) around a quasar that emits ultraviolet photons into a partially neutral IGM. Before the overlap phase of ionized hydrogen (H II) regions that signals the end of the reionization epoch⁹, galaxies generated their own small ionized regions, filling a fraction of the IGM volume around the quasar. We define the neutral fraction, $x_{\text{H I}}$, to be the mean fraction of hydrogen atoms (in a representative volume of the low-density IGM) that remain neutral before the quasar activity, and assume that $x_{\text{H I}}$ outside the Strömgren sphere remains constant during its expansion. Because the luminosity of a bright quasar is much larger than the stellar luminosity of neighbouring galaxies, a fully ionized quasar bubble expands into a partially ionized IGM composed of isolated H II regions.

Even for the deepest available spectra, with lower limits on the Ly α (Gunn–Peterson¹⁰) optical depth of $\tau_{\text{GP}} \gtrsim 26$, the IGM need only have a neutral fraction of $\sim 10^{-3}$ to produce the observed absorption blueward of Ly α ¹. However, the value of $R_{\text{obs}} \approx 4.5 \text{ Mpc}$ may be used to constrain larger values of the neutral fraction in

the range $0.001 \leq x_{\text{H I}} \leq 1.0$. In the early expansion phase, the dependence of the physical radius of the Strömgren sphere, R_p , on the quasar age^{1,11}, t_{age} , may be crudely approximated as $R_p \approx 7x_{\text{H I}}^{-1/3} (t_{\text{age}}/10^7 \text{ yr})^{1/3} \text{ Mpc}$, given the production rate of ionizing photons that is characteristic of SDSS J1148+5251 and SDSS J1030+0524. Clearly, the Strömgren sphere grows larger when embedded in an IGM with a small neutral fraction. Defining R_{max} to be the maximum radius achieved in a fully neutral IGM by the end of the quasar lifetime, t_{lt} , the resulting a priori probability distribution is $dP/dR_p|_{x_{\text{H I}}} = 3x_{\text{H I}}R_p^2/R_{\text{max}}^3$ for $R_p < R_{\text{max}}x_{\text{H I}}^{-1/3}$. In addition, the neutral fraction must be smaller than $x_{\text{max}} = \min([t_{\text{lt}}/10^7 \text{ years}] [R_{\text{obs}}/7 \text{ Mpc}]^{-3}, 1)$ to allow the growth of the sphere to R_{obs} within t_{lt} . Using a prior probability distribution that is flat in the logarithm of $x_{\text{H I}}$, we find the cumulative a posteriori probability distribution, $P(<x_{\text{H I}}) = \min(x_{\text{H I}}, x_{\text{max}})/x_{\text{max}}$. Because estimates of quasar lifetimes¹² are bracketed in the range 10^6 – 10^8 years, we get $x_{\text{max}} \approx 1$ and conclude that $x_{\text{H I}} \gtrsim 0.1$ (0.01) with 90% (99%) confidence.

To quantify this simple argument better, we write the general equation for the relativistic expansion of the co-moving radius [$R = (1+z)R_p$] of the quasar H II region¹ in an IGM with a neutral filling fraction $x_{\text{H I}}$ (fixed by other ionizing sources):

$$\frac{dR}{dt} = c(1+z) \left[\frac{F_{\gamma} \dot{N}_{\text{ion}} - \alpha_B C F_m x_{\text{H I}} (\bar{n}_0^{\text{H}})^2 (1+z)^3 \left(\frac{4\pi}{3} R^3\right)}{F_{\gamma} \dot{N}_{\text{ion}} + 4\pi R^2 (1+z) c F_m x_{\text{H I}} \bar{n}_0^{\text{H}}} \right] \quad (1)$$

where c is the speed of light, $\bar{n}_0^{\text{H}}$ is the mean number density of protons at $z = 0$, $\alpha_B = 2.6 \times 10^{-13} \text{ cm}^3 \text{ s}^{-1}$ is the case-B recombination coefficient at the characteristic temperature of 10^4 K , and $\dot{N}_{\text{ion}}$ is the rate of ionizing photons crossing a shell at the radius of the H II region at time t . We use the distribution derived from numerical simulations for the over-densities in gas clumps¹³, and calculate the mean free path $d(\Delta_c)$ for ionizing photons^{13,14}. Following ref. 14, we then find the critical overdensity (Δ_c) at which a fraction $F_{\gamma} \equiv 0.5 = \exp[-R_p/d(\Delta_c)]$, of photons emitted do not encounter an overdensity larger than Δ_c within the H II region. We also compute the mass fraction $F_m (\sim 1)$ of gas within R_p that is at overdensities lower than Δ_c . Finally, we calculate the clumping factor in the ionized regions, $C(R) \equiv \langle \Delta^2 \rangle / \langle \Delta \rangle^2$, where the angular brackets denote an average over all regions with $\Delta < \Delta_c$. For $x_{\text{H I}} = 1$ and $dR_p/dt \ll c$, equation (1) reduces to its well-known form^{11,15–17}. The expansion of the quasar Strömgren sphere would change in an overdense region of the IGM. However, the density contrast due to infall has been shown¹⁴ to be small ($\sim 1\%$) on scales of several megaparsecs around the massive hosts of the bright SDSS quasars at $z \approx 6$. Numerical experiments showed our results to be insensitive to infall. The emission rate of ionizing photons $\dot{N}_{\text{ion}}$ in equation (1) is computed at $t' = t - t_{\text{delay}}$ to account for the finite light travel time between the source and the ionization front. The delay t_{delay} is derived from the relation $R = \int_{t_{\text{R}}-t_{\text{delay}}}^{t_{\text{R}}} c dt' [1+z(t')]$, where t_{R} is the time when the photon crosses the co-moving radius R . The use of the Telfer *et al.*¹⁸ spectral template implies $\dot{N}_{\text{ion}} = 6.5 \times 10^{57} \text{ s}^{-1}$ for SDSS J1030+0524 and $\dot{N}_{\text{ion}} = 10.0 \times 10^{57} \text{ s}^{-1}$ for SDSS J1148+5251, whereas the template of Elvis *et al.*¹⁹ implies lower values of $\dot{N}_{\text{ion}} \approx 1.3 \times 10^{57} \text{ s}^{-1}$ and $\dot{N}_{\text{ion}} \approx 2.0 \times 10^{57} \text{ s}^{-1}$, respectively¹.

Our model for the evolution of the Strömgren sphere includes the quasar formation history. We compute the time-dependent ionizing flux governed by black-hole growth through accretion and mergers. Following the observational inference that the relation between bulge velocity dispersion and black-hole mass M_{bh} does not evolve with redshift²⁰, we extrapolate the present-day $M_{\text{bh}}-M_{\text{halo}}$ relation²¹ using the dependence of virial velocity on redshift⁹ to obtain $M_{\text{halo}} = 1.5 \times 10^{12} M_{\odot} (M_{\text{bh}}/10^9 M_{\odot})^{3/5} [(1+z)/7]^{-3/2}$. The luminosities of the SDSS quasars imply black-hole masses of $\sim 2 \times 10^9 M_{\odot}$ accreting at their Eddington rate⁶, leading to inferred host dark-matter halos of $M_{\text{halo}} \approx 2 \times 10^{12} M_{\odot}$. The inference of such a

massive halo at this early epoch is supported by the signature of gas infall in the spectra of high redshift quasars¹⁴, and the inference of a large molecular mass and velocity width in the host galaxy of SDSS J1148+5251 (ref. 7).

We begin with two dark-matter haloes of comparable mass, M_1 and M_2 , that merge to form the host dark-matter halo with a mass $M_{\text{halo}} = (M_1 + M_2)$ at a redshift corresponding to one quasar lifetime before the observed redshift, and construct merger trees for each of the progenitor haloes using the algorithm described in ref. 22. We use the $M_{\text{bh}}-M_{\text{halo}}$ relation to find the mass of the black hole at the centre of each dark-matter halo in the merger tree. When there is a major merger (having a progenitor mass ratio < 2), we assume that the black holes merge and that the accretion of the mass deficit relative to the $M_{\text{bh}}-M_{\text{halo}}$ relation produces the limiting Eddington luminosity with a radiative efficiency of $\epsilon = 10\%$. The resulting quasar lifetime is $t_{\text{lt}} = 4 \times 10^7 (\epsilon/0.1) \ln[M_{\text{halo}}^{5/3}/(M_1^{5/3} + M_2^{5/3})]$ years. The ionizing photon emission rate $\dot{N}(t)$ is taken to have an exponential time dependence, $\dot{N}_{\text{ion}}(t) = \dot{N}_0 e^{-t/t_{\text{lt}}}$, with $\dot{N}_0 \propto M_{\text{bh}}$. We assume that all quasar episodes in the merger tree contribute to the volume of the observed Strömgren sphere, which is centred on the most massive halo in the tree. A similar prescription for quasar activity was shown to be consistent with the observed number counts of high redshift quasars²². For a major

merger resulting in the observed quasar activity, the lifetime given by this approach is $\sim 10^7$ years. This compares favourably with a variety of observational estimates of quasar lifetimes^{23–26}. To cover the full range of uncertainty we consider modifications to our fiducial model in which the quasar lifetime is multiplied a factor $f_{\text{lt}} = 0.1$ – 10 , resulting in lifetimes for the observed quasars of $f_{\text{lt}} t_{\text{lt}} = 10^6$ – 10^8 years.

We have produced 300 realizations of the merger tree, and computed the evolution of the Strömgren sphere in each case for the full range of neutral fractions $0.001 \leq x_{\text{H I}} \leq 1.0$ that is allowed by the Gunn–Peterson optical depth¹. We find that quasar activity associated with the hierarchical build-up of the host galaxy produces an H II region with a typical size of $\sim 2x_{\text{H I}}^{-1/3}$ Mpc, comparable to the observed radii if $x_{\text{H I}}$ is of order unity. Figure 1 shows the conditional a priori probability distributions of R_p for quasars like SDSS J1030+0524, assuming $x_{\text{H I}} = 1$ and two different values of f_{lt} . The observed radii are consistent with lifetimes close to the fiducial case of $f_{\text{lt}} = 1$. The plotted distribution of R_p implies that the hierarchical evolution of early quasars leads to a dearth of small

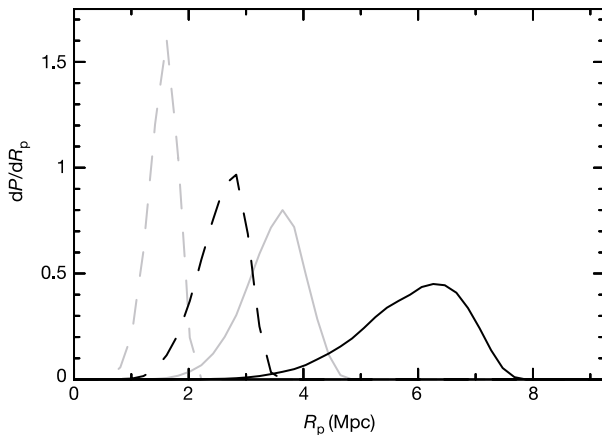


Figure 1 Predicted probability for observing different radii of the ionized region around the quasar SDSS J1030+0524. The differential a priori probability distribution is plotted for $x_{\text{H I}} = 1$ and two quasar lifetimes: $f_{\text{lt}} = 0.1$ (dashed) and $f_{\text{lt}} = 1.0$ (solid). The light and dark curves show results calculated from the Elvis *et al.*¹⁹ and Telfer *et al.*¹⁸ template spectra, respectively. The distributions may also be used to approximate the behaviour with $x_{\text{H I}} < 1$ by replacing R_p with $R_p x_{\text{H I}}^{-1/3}$. Values of R_p significantly larger than observed still allow transmission of Ly α flux at a detectable level because: (1) the optical depth due to the damping wing of the IGM²⁹ is only important near the boundary of the H II region, and (2) the optical depth at R_p due to resonant absorption within the H II region¹⁴ is in the range $\tau_{\text{reson}} = (1-8) \times (R_p/4.5 \text{ Mpc})^2$. A dense H I cloud near R_{obs} could produce a large damping wing and lead to a systematic underestimation of R_p . However, such a cloud would need a column density $\geq 10^{21} \text{ cm}^{-2}$ ($\geq 10^{22} \text{ cm}^{-2}$) to produce a Ly α optical depth > 26 over 10% (30%) of the spectral range covered by the H II region. This requires a damped Ly α absorber, whose existence is highly improbable within the narrow redshift interval under consideration, $\Delta z \approx 0.1$. Our calculation assumes that the ionization front is thin. The spectrum-averaged mean-free path for the ionizing quasar photons is $\lambda \approx 1.5x_{\text{H I}}^{-1}(1+z)^{-3} \text{ Mpc}$, which may be compared to the bubble radius $R_p \approx 4.5x_{\text{H I}}^{-1/3} \text{ Mpc}$, to yield the fractional thickness of the ionization front $(\lambda/R_p) \approx 10^{-3}(R_p/4.5 \text{ Mpc})^{-1}x_{\text{H I}}^{-2/3}[(1+z)/7.3]^{-3}$. We also note that the ionized region need not be spherical; if the quasar radiation is beamed, its luminosity per unit solid-angle along the observer's line-of-sight is still measured, and equation (1) therefore describes the extent of the ionized region along the line-of-sight. Throughout our calculations, we have adopted the best-fit cosmological parameters derived from the WMAP data³⁰.

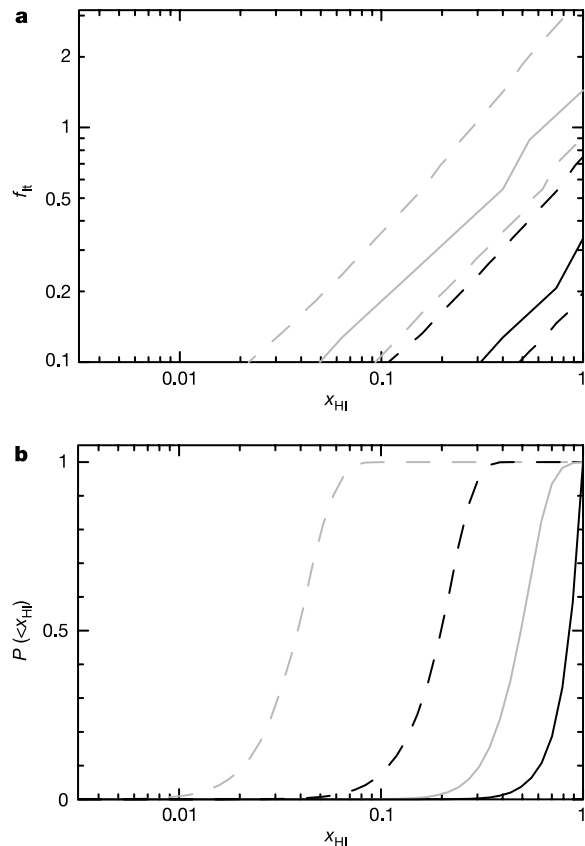


Figure 2 Likelihood for the inferred neutral fraction of the IGM, assuming different quasar lifetimes. We show contours of likelihood, L , per logarithm of $x_{\text{H I}}$ and f_{lt} (a) and a posteriori cumulative probability distributions for $x_{\text{H I}}$ (b). Cumulative distributions are shown for two different quasar lifetimes: $f_{\text{lt}} = 0.1$ (dashed) and $f_{\text{lt}} = 1.0$ (solid). The light and dark lines show results where the quasar ionizing photon rate was specified by the Elvis *et al.*¹⁹ and Telfer *et al.*¹⁸ spectra, respectively. The a posteriori probability is $\frac{dP}{dx_{\text{H I}}} \propto \left[\frac{dP_{1030}}{dR_p}(R_p = 4.5 \text{ Mpc}|x_{\text{H I}}) \times \frac{dP_{1148}}{dR_p}(R_p = 4.7 \text{ Mpc}|x_{\text{H I}}) \right] \frac{dP_{\text{prior}}}{dx_{\text{H I}}}$, where $\frac{dP_{\text{prior}}}{dx_{\text{H I}}}$ is the prior probability for $x_{\text{H I}}$, assumed to be flat in logarithmic bins within the range $0.001 \leq x_{\text{H I}} \leq 1.0$, and $\frac{dP_{1030}}{dR_p}$ and $\frac{dP_{1148}}{dR_p}$ are the a priori probability distributions for R_p in quasars like SDSS J1030+0524 and SDSS J1148+5251, respectively. A flat logarithmic prior is the natural choice for $\frac{dP_{\text{prior}}}{dx_{\text{H I}}}$, because $x_{\text{H I}}$ may be thought of as a ratio of independent quantities (the number of ionizing photons and the number of baryons) that would themselves have linearly distributed prior probabilities. The alternative use of a linear prior in $x_{\text{H I}}$ leads to more stringent limits than those presented here.

Strömgren sphere radii for the observed SDSS quasars. This is in contrast to the simplified monolithic model of quasar formation^{16,1}, for which the H II region has zero volume at the time when the quasar turns on [$R(t_{\text{age}} = 0) = 0$] and the distribution extends down to $R_p = 0$.

To constrain the neutral fraction, we have computed the likelihood of observing $R_{\text{obs}} = 4.5$ Mpc around SDSS J1030+0524 and $R_{\text{obs}} = 4.7$ Mpc around SDSS J1148+5251 as a function of x_{HI} and f_{lt} . The results are plotted in Fig. 2. Figure 2a shows the locus of most likely values (thick line) as well as likelihood contours at 0.1 of the peak value (dashed lines). The fiducial lifetime (f_{lt}) favours $x_{\text{HI}} \approx 1$, while with $f_{\text{lt}} \ll 1$ the distributions for R_p lie substantially below the observed value of 4.5 Mpc, making smaller values of x_{HI} more likely. Extrapolation of the most likely contour for the Elvis *et al.*¹⁹ template yields $f_{\text{lt}} \approx 2x_{\text{HI}}$. This implies that $x_{\text{HI}} \approx 10^{-3}$ would require a lifetime as short as 2×10^4 years, which is ruled out by the variability properties of quasars in SDSS²⁶. The a posteriori probability distributions for x_{HI} given f_{lt} are plotted in Fig. 2b and robustly yield the constraint $x_{\text{HI}} \geq 0.01$. For $f_{\text{lt}} \geq 0.3$, we find $x_{\text{HI}} \geq 0.1$ and $x_{\text{HI}} \geq 0.4$ at the 90% level for the Elvis *et al.*¹⁹ and Telfer *et al.*¹⁸ spectra, respectively. The fiducial model with $f_{\text{lt}} = 1$ yields corresponding constraints of $x_{\text{HI}} \geq 0.3$ and $x_{\text{HI}} \geq 0.6$.

The inference of a large neutral fraction at $z \approx 6.3$ presents a challenge to theories of cosmological reionization when combined with the large optical depth³ to electron scattering after cosmological recombination, $\tau_{\text{es}} = 0.17 \pm 0.04$. Consider a toy model in which the Universe was partially reionized at a high redshift $z_{\text{reion,I}}$, leaving a neutral fraction x_{HI} until complete reionization was reached at $z_{\text{reion,II}} \approx 6.25$. The optical depth is then $\tau_{\text{es}} = 0.04 + 0.002(1 - x_{\text{HI}})[(1 + z_{\text{reion,I}})^{3/2} - 19.5]$. If the IGM ionization fraction increased monotonically, then given $\tau_{\text{es}} > 0.13$, the Universe needed to be reionized earlier than $z_{\text{reion,I}} \approx 18$ or $z_{\text{reion,I}} \approx 24$, assuming the Elvis *et al.*¹⁹ and Telfer *et al.*¹⁸ spectra, respectively. If $\tau_{\text{es}} = 0.17$ and $x_{\text{HI}} > 0.7$, as implied by the Telfer *et al.*¹⁸ spectrum, then a monotonic reionization history requires significant reionization at an implausibly high redshift ($z \geq 30$). In this case, a more plausible alternative is a non-monotonic history with an early reionization peak, possibly due to the formation of massive population-III stars^{27,28}. □

Received 14 October 2003; accepted 15 January 2004; doi:10.1038/nature02336.

- White, R. L., Becker, R. H., Fan, X. & Strauss, M. A. Probing the ionization state of the universe at $z > 6$. *Astron. J.* **126**, 1–14 (2003).
- Fan, X. *et al.* Evolution of the ionizing background and the epoch of reionization from the spectra of $z \sim 6$ quasars. *Astron. J.* **123**, 1247–1257 (2002).
- Kogut, A. *et al.* First-year Wilkinson microwave anisotropy probe (WMAP) observations: temperature-polarization correlation. *Astron. J. Suppl.* **148**, 161–173 (2003).
- Hu, E. M. *et al.* A redshift $z = 6.56$ galaxy behind the cluster Abel 370. *Astrophys. J.* **568**, L75–L79 (2002).
- Haiman, Z. The detectability of high-redshift Ly α emission lines prior to the reionization of the universe. *Astrophys. J.* **576**, L1–L4 (2002).
- Willott, C. J., McLure, R. J. & Jarvis, M. J. A $3 \times 10^9 M_{\odot}$ black hole in the quasar SDSS J1148+5251 at $z = 6.41$. *Astrophys. J.* **587**, L15–L18 (2003).
- Walter, F. *et al.* Molecular gas in the host galaxy of a quasar at redshift $z = 6.42$. *Nature* **424**, 406–408 (2003).
- Richards, G. T. *et al.* Broad emission-line shifts in quasars: an orientation measure for radio-quiet quasars? *Astron. J. Suppl.* **124**, 1–17 (2002).
- Barkana, R. & Loeb, A. In the beginning: the first sources of light and the reionization of the universe. *Phys. Rep.* **349**, 125–238 (2001).
- Gunn, J. E. & Peterson, B. A. On the density of neutral hydrogen in intergalactic space. *Astrophys. J.* **142**, 1633–1641 (1965).
- Pentericci, L. *et al.* VLT optical and near-infrared observations of the $z = 6.28$ quasar SDSS J1030+0524. *Astron. J.* **123**, 2151–2158 (2002).
- Martini, P. QSO lifetimes. Preprint at (<http://arXiv.org/astro-ph/0304009>) (2003).
- Miralda-Escude, J., Haehnelt, M. & Rees, M. Reionization of the inhomogeneous universe. *Astrophys. J.* **530**, 1–16 (2000).
- Barkana, R. & Loeb, A. GRBs versus Quasars: Lyman- α signatures of reionization versus cosmological infall. *Astrophys. J.* (in the press); preprint at (<http://arXiv.org/astro-ph/0305470>) (2003).
- Shapiro, P. R. & Giroux, M. L. Cosmological, H II regions and the photoionization of the intergalactic medium. *Astrophys. J.* **321**, L107–L112 (1987).
- Cen, R. & Haiman, Z. Quasar Stromgren spheres before cosmological reionization. *Astrophys. J.* **542**, L74–L78 (2000).
- Madau, P. & Rees, M. J. The earliest luminous sources and the damping wing of the Gunn-Peterson trough. *Astrophys. J.* **542**, L69–L73 (2000).
- Telfer, R. C., Zheng, W., Kriss, G. A. & Davidsen, A. F. The rest-frame extreme-ultraviolet spectral properties of quasi-stellar objects. *Astron. J.* **565**, 773–785 (2002).
- Elvis, M. *et al.* Atlas of quasar energy distributions. *Astron. J. Suppl.* **95**, 1–68 (1994).
- Shields, G. A. *et al.* The black hole-bulge relationship in quasars. *Astrophys. J.* **583**, 124–133 (2003).
- Ferrarese, L. Beyond the Bulge: a fundamental relationship between supermassive black holes and dark matter halos. *Astrophys. J.* **587**, 90–97 (2001).
- Volonteri, M., Haardt, F. & Madau, P. The assembly and merging history of supermassive black holes in hierarchical models of galaxy formation. *Astrophys. J.* **582**, 559–573 (2003).
- Yu, Q. & Tremaine, S. Observational constraints on growth of massive black holes. *Mon. Not. R. Astron. Soc.* **335**, 965–976 (2002).
- Martini, P. & Weinberg, D. H. Quasar clustering and the lifetime of quasars. *Astrophys. J.* **547**, 12–26 (2001).
- Jakobsen, P., Jansen, R. A., Wagner, S. & Reimers, D. Caught in the act: a helium-reionizing quasar near the line of sight to Q0302003. *Astron. Astrophys.* **397**, 891–898 (2003).
- Martini, P. & Schneider, D. P. Preprint at (<http://arXiv.org/astro-ph/0309650>) (2003).
- Wyithe, J. S. B. & Loeb, A. Reionization of hydrogen and helium by early stars and quasars. *Astrophys. J.* **586**, 693–708 (2003).
- Cen, R. The implications of Wilkinson microwave anisotropy probe observations for population III star formation processes. *Astrophys. J.* **591**, L5–L8 (2003).
- Miralda-Escude, J. Reionization of the intergalactic medium and the damping wing of the Gunn-Peterson trough. *Astrophys. J.* **501**, 15–22 (1998).
- Spergel, D. N. *et al.* First-year Wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters. *Astron. J. Suppl.* **148**, 175–194 (2003).

Acknowledgements This work was supported in part by grants from ARC, NSF and NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.L. (aloeb@cfa.harvard.edu).

Atomic transient recorder

R. Kienberger^{1*}, E. Goulielmakis^{1*}, M. Uiberacker^{1*}, A. Baltuska¹, V. Yakovlev¹, F. Bammer², A. Scrinzi¹, Th. Westerwalbesloh³, U. Kleineberg³, U. Heinzmann³, M. Drescher³ & F. Krausz^{1,4}

¹Institut für Photonik, Technische Universität Wien, Gusshausstraße 27, A-1040 Wien, Austria

²Institut für Spanlose Fertigung und Hochleistungslasertechnik, Technische Universität Wien, Franz-Grillstr. 1, Arsenal Obj. 207, A-1030 Wien, Austria

³Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany

⁴Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany

*These authors contributed equally to this work

In Bohr's model of the hydrogen atom, the electron takes about 150 attoseconds ($1 \text{ as} = 10^{-18} \text{ s}$) to orbit around the proton, defining the characteristic timescale for dynamics in the electronic shell of atoms. Recording atomic transients in real time requires excitation and probing on this scale. The recent observation of single sub-femtosecond ($1 \text{ fs} = 10^{-15} \text{ s}$) extreme ultraviolet (XUV) light pulses¹ has stimulated the extension of techniques of femtochemistry² into the attosecond regime^{3,4}. Here we demonstrate the generation and measurement of single 250-attosecond XUV pulses. We use these pulses to excite atoms, which in turn emit electrons. An intense, waveform-controlled, few cycle laser pulse⁵ obtains 'tomographic images' of the time-momentum distribution of the ejected electrons. Tomographic images of primary (photo)electrons yield accurate information of the duration and frequency sweep of the excitation pulse, whereas the same measurements on secondary (Auger) electrons will provide insight into the relaxation dynamics of the electronic shell following excitation. With the current $\sim 750\text{-nm}$ laser probe and $\sim 100\text{-eV}$ excitation, our transient recorder is capable of resolving atomic electron dynamics within the Bohr orbit time.

In electron-optical chronoscopy the rapid deflection of a bunch

of photoelectrons maps their temporal profile to a spatial distribution^{6,7}. From the electrons' streaked image the temporal structure of the light pulse triggering the photoemission can be inferred with sub-picosecond resolution. The resolution of streak cameras is limited by the spread of the electron transit time owing to the spread of their initial momenta. In this work we report sampling of sub-femtosecond electron emission from atoms by drawing on the same basic concept. The improved time resolution results from several essential modifications of image-tube streak cameras: the oscillating electric field of light is used for affecting the electrons' motion (1), the field is virtually jitter-free (2), and is applied along the direction of electron motion, which implies their acceleration or deceleration instead of their deflection (3), directly at the location and instant of emission (4). Whereas (1) implies 'only' a dramatically enhanced streaking speed, the consequences of (2)–(4) are more profound: (2) allows us to systematically vary the timing of the probing field with an accuracy within the electron bunch length, and because of (3) this capability results in the 'projection' of the initial time–momentum distribution of electron emission into a series of distributions of the electrons' final momentum, while (4) prevents the initial momentum spread from introducing any measurement error. As a result of (2)–(4), the spread of initial electron momenta no longer limits the time resolution. Moreover, a possible temporal variation of this initial momentum spread during the emission can be captured just as can the temporal variation of the emission intensity.

The final momentum distribution of electrons detached from atoms by an impulsive excitation in the presence of an intense,

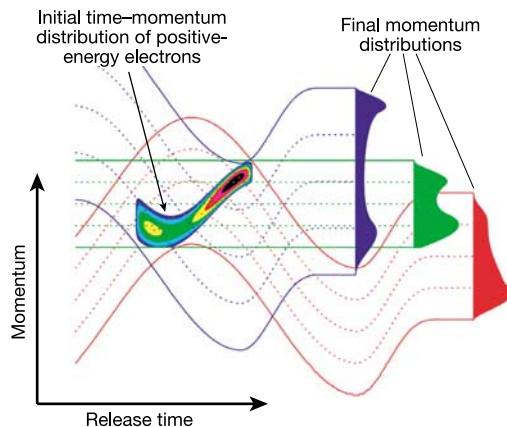


Figure 1 Principle of the atomic transient recorder. The initial time–momentum distribution of positive-energy (photo or Auger) electrons emitted from atoms excited by an attosecond pulse carries direct time-domain information about the excitation and relaxation dynamics of the electronic shell. A linearly polarized, few-cycle laser pulse, with its electric field directed parallel to the direction of observation of the ejected electrons, creates at different delays different tomographic projections of the time–momentum distribution $n_e(p_i, t)$ of atomic electron emission on momentum space. These final momentum distributions can readily be measured after the laser pulse left the interaction region. In the absence of the laser field, the electron momenta do not change after detachment and accumulation of electrons with a given final momentum (mathematically: time-integration) follows lines of constant momentum (green lines) yielding the field-free final momentum distribution: $\sigma_0(p) = \int_{-\infty}^{\infty} n_e(p, t) dt$ (green profile). In the presence of a probing field, accumulation of electrons with a given final momentum occurs along the lines of constant canonical momentum $p_i + eA_L(t)$ (blue and red lines, see equation (1)). The corresponding projections (streaked spectra) are represented in red and blue. From a suitable set of such tomographic projections the time–momentum distribution of electron emission can be retrieved, providing direct time-domain insight into atomic dynamics triggered by an attosecond excitation (in our experiments XUV pulse) synchronized to the probing laser field.

linearly polarized laser field $E_L(t) = E_0(t)\cos(\omega_L t + \varphi)$ and observed within a narrow cone aligned parallel to the laser electric field vector is given by:

$$\sigma(p) = \int_{-\infty}^{\infty} n_e(p - eA_L(t), t) dt \quad (1)$$

where $n_e(p_i, t)$ is the initial time–momentum distribution of emitted electrons and e is the electron charge. The vector potential in the Coulomb gauge $A_L(t) = \int_t^{\infty} E_L(t') dt'$ accounts for the

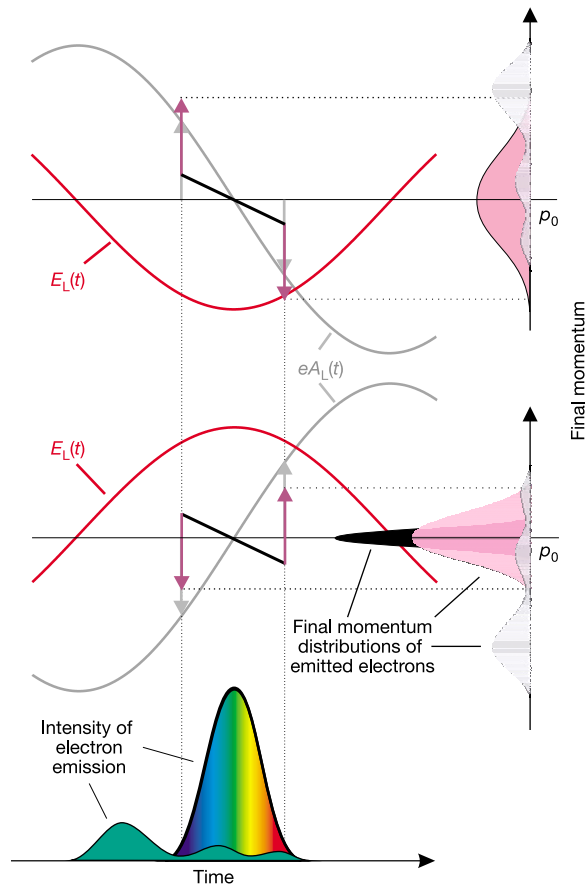


Figure 2 Electron streak records in specific cases. The electric field vector of the linearly polarized probe laser pulse points towards the electron detector for $E_L > 0$. In the absence of a temporal sweep of initial momenta of the emitted atomic electrons, described by $n_e(p_i, t) = f(p_i)g(t)$, the momentum transfer $\Delta p = eA_L(t_i)$ (grey line), maps the temporal emission profile $g(t)$ (green profile) uniquely into a similar distribution of final momenta (grey profiles) around the mean initial momentum p_0 , provided that the emission is temporally confined between adjacent zero transitions of the probing laser field $E_L(t)$. In this specific case, the temporal intensity profile of electron emission is directly mirrored by the streak record just as in a conventional streak camera. However, any sweep of the electrons' initial momentum (for example, a linear one as indicated by the black line for the single-peaked electron emission) revokes the unique correspondence between an electron's final momentum and its release time, preventing the retrieval of accurate temporal information from single streak records. In this case, at least two streak records (pink profiles)—in addition to the field-free spectrum (black profile)—are required. In the absence of a nonlinear temporal momentum sweep the streak records obtained near the zero transitions of $A_L(t)$ with opposite slopes (pink profiles) together with the field-free spectrum (black profile) allow determination of all relevant characteristics: the temporal profile, duration and momentum chirp of emission. Note that a linear momentum sweep leads to an asymmetric broadening of the streaked spectra at these delays and it is this asymmetry that makes the measurement highly sensitive to the momentum sweep, that is, highly sensitive to deviations of the pulse duration from the Fourier limit.

momentum boost imparted to the free electron by the laser field from the time of its release t until the pulse has passed. In the classical description of the freed electrons' motion in the strong laser field, the final momentum spectrum can be viewed as a projection of $n_e(p_i, t)$ along the lines where $p = p_i + eA_L(t)$ is constant (see Fig. 1); p_i is the initial kinetic momentum of the electron. By delaying the laser field with respect to the excitation that triggers electron emission, a set of tomographic records is obtained, from which the initial time-momentum distribution $n_e(p_i, t)$ can be uniquely reconstructed if the emission terminates within $T_0/2 = \pi/\omega_L$. The method is closely related to frequency-resolved optical gating^{8,9} with the oscillating laser field (rather than its envelope) being employed as a gate. In the simplest cases one or

two streaked spectra are sufficient for retrieving $n_e(p_i, t)$, as discussed in Fig. 2.

For our experiments the XUV excitation pulses have been generated from neon atoms ionized by an intense linearly polarized few-cycle laser pulse (full-width at intensity half-maximum (FWHM): $\tau_L = 5$ fs, carrier wavelength: $\lambda_L = 750$ nm, $T_0 = 2.5$ fs) with controlled waveform⁵ by a process that leads to the emission of high-order harmonics for multi-cycle drivers^{10,11}. Proper adjustment of the laser peak intensity yielded highest-energy (cutoff) XUV emission in the high-reflectivity band of our Mo/Si multilayer mirror, which was used for focusing the XUV as well as laser pulses into another neon gas jet for recording atomic transients. Both an intuitive model^{12,13} and extensive numerical simulations^{14–18} predict the cutoff emission to be confined to the vicinity of zero transition(s) of the laser electric field after the most intense half-cycle(s) in a few-cycle driver.

Figure 3 summarizes representative streaked neon photoelectron spectra recorded with the XUV and laser pulse impinging with a fixed relative timing set in the XUV generation process. For a cosine driver waveform ($\varphi = 0$), cutoff radiation (filtered by the Mo/Si multilayer) is predicted to be emitted in a single bunch at the zero transition of $E_L(t)$ following the pulse peak. The photoelectrons knocked off in the direction of the peak electric field at this instant should gain the maximum increase of their momentum and energy. Figure 3d corroborates this prediction. The clear upshift is consistent with the XUV burst coinciding with the zero transition of the laser electric field (see Fig. 2). Possible satellites would appear at the adjacent zero transitions of $E_L(t)$ and suffer an energy downshift. The absence of a downshifted spectral peak of substantial intensity indicates a clean single sub-fs pulse generation. With the phase

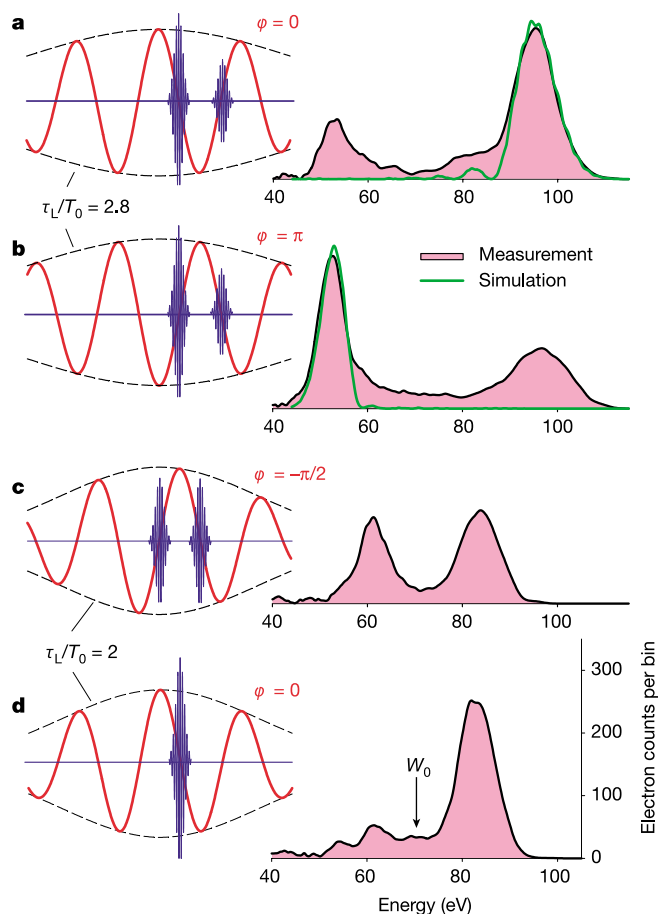


Figure 3 Streaked photoelectron spectra recorded at a fixed delay of probe laser light. Energy distribution of photoelectrons emitted from neon atoms excited by a sub-fs XUV pulse carried at a photon energy of $\hbar\omega_{XUV} \approx 93.5$ eV (selected by the Mo/Si mirror). The photoelectron spectrum peaks at $W_0 = \hbar\omega_{XUV} - W_b \approx 72$ eV in the absence of $E_L(t)$, where $W_b = 21.5$ eV is the binding energy of the most weakly bound valence electrons in Ne. The spectrally filtered cutoff XUV bursts and the 5-fs, 750-nm driver laser pulses are depicted by blue and red lines, respectively. **a, b**, Streaked spectra obtained with respectively 'cosine' and '-cosine' laser pulses of a normalized duration of $\tau_L/T_0 = 2.8$ and of a peak electric field of $E_0 = 140$ MV cm⁻¹. The green lines on the right-hand side depict spectra computed with an XUV burst derived from the measured asymmetric XUV radiation filtered by the mirror (see Fig. 5a) under the assumption of zero spectral phase. The satellite pulse is not modelled in this way, because the corresponding modulation of the spectrum was not considered in the calculation. The difference in broadening of the up- and down-shifted spectral features appears to be a consequence of the quadratic temporal frequency sweep resulting from the asymmetric spectral distribution of the XUV burst (see Fig. 5a). **c, d**, Streaked spectra obtained with respectively 'sine' and 'cosine' laser pulses characterized by $\tau_L/T_0 = 2$ and $E_0 = 75$ MV cm⁻¹.

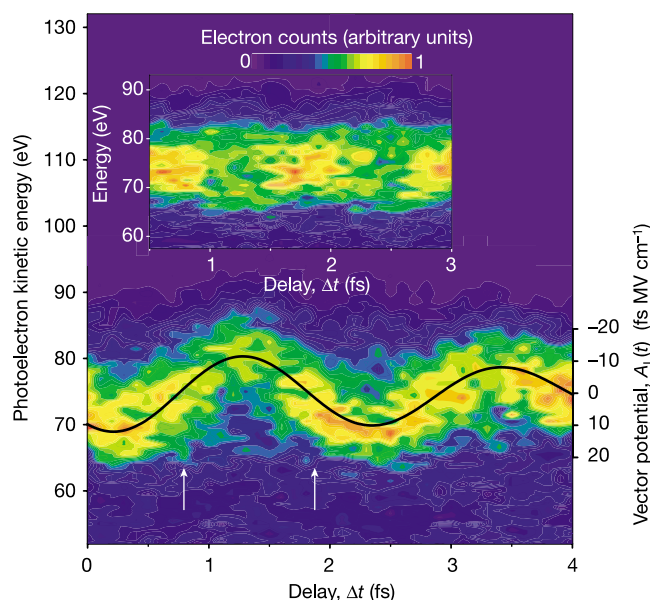


Figure 4 ATR measurement: a series of tomographic projections (streaked kinetic energy spectra) of the initial time-momentum distribution of photoelectrons knocked out by a single sub-fs XUV pulse (in false-colour representation). A few-cycle laser pulse with a cosine waveform and a normalized duration of $\tau_L/T_0 = 2$ was used for both generating the single 93-eV sub-fs excitation pulse and for probing photoelectron emission in the atomic transient recorder. Black line, $A_L(t)$ of the probing field evaluated from the peak shift of the streaked spectra (see scale on the right hand side). From $A_L(t)$ the electric field of the light wave can also be determined by using the relationship $E_L(t) = -dA_L/dt$. The electric field vector points towards the electron detector for $E_L > 0$. Inset, streaked spectra obtained under the same experimental conditions except for the carrier-envelope phase of the few-cycle pulse, which was left unstabilized.

adjusted to yield a sine waveform ($\varphi = -\pi/2$), cutoff emission is predicted to come in twin pulses (Fig. 3c). The double-peaked streaked spectrum (Fig. 3c) clearly reflects this time structure. High-energy XUV photons are now distributed in two bursts, each of which is less than half as intense as the isolated burst produced by the cosine waveform (Fig. 3d). These measurements demonstrate how light waveform control allows shaping XUV emission on a sub-fs timescale.

Other than φ the normalized pulse duration τ_L/T_0 plays a crucial role in the generation of a single XUV pulse, as demonstrated by experiments with slightly longer, 7-fs laser pulses ($\tau_L/T_0 = 2.8$). Under these conditions a satellite pulse is unavoidable for any setting of φ and most careful filtering of the cutoff photons (Fig. 3a, b). A reduced difference in intensity of the adjacent half-cycles of the few-cycle wave implies that the highest-energy spectral components of adjacent XUV bursts reach the energy band selected by our 15%-bandwidth bandpass filter.

With isolated sub-fs XUV pulses at our disposal atomic transients can now be triggered and their subsequent evolution be captured by probing electron emission with a synchronized wave of laser light in the atomic transient recorder (ATR). In the first ATR measurements presented here the objects of scrutiny were photoelectrons. Figure 4

shows a series of streaked spectra of photoelectrons emitted from neon at different delays of $E_L(t)$. If the electrons' initial kinetic energy is much larger than their average quiver energy, their laser-induced energy shift $\Delta W(t) \approx (p_0/m)\Delta p(t) = (ep_0/m)A(t)$, where m is the electron mass, probes $A_L(t)$ directly at the instant t of electron release and allows its determination without having to analyse the structure of the streaked spectra. The evaluated $A_L(t)$ is depicted by the black line in Fig. 4, revealing $T_0/2 \approx 1$ fs at the pulse peak, which is indicative of a strong blue shift. This was observed previously and found to originate from ionization-induced self-phase modulation in the XUV source¹. With $A_L(t)$ known, we can set out to characterize the electron emission. Waveform control is as essential here as for the reproducible generation of sub-fs bursts (see Fig. 3); its absence smears the streak records beyond redemption (inset in Fig. 4).

Figure 5 depicts the photoelectron spectra in the absence of the probe laser field (Fig. 5a) and the streaked spectra (Fig. 5b, c) recorded at adjacent zero transitions of $A_L(t)$, which are most sensitive to a possible quadratic spectral phase (linear chirp) carried by the XUV pulse. The best fits to these data (blue lines in Fig. 5) yield $\tau_{\text{XUV}} = 250(-5/+30)$ as and $\tau_{\text{jitter}} = 260(\pm 80)$ as for the duration and jitter of the XUV pulse, respectively. The remarkable accuracy of τ_{XUV} derives from the use of several tomographic projections of the time-momentum distribution of photoelectrons (spectra shown in Fig. 5). The inspection of other streaked spectra justified the neglect of higher-order spectral phase in the above analysis. The large jitter appears to be extrinsic (vibrations due to the vacuum pumps) because it is virtually absent in the absence of a delay between XUV and laser pulse (Fig. 3). The temporal intensity profile and chirp of the XUV pulse obtained from the ATR measurements are shown in Fig. 5d. The pulse is Fourier-limited with a weak temporal chirp resulting from the asymmetric shape of the pulse spectrum (Fig. 5a).

The briefest time interval within which two different atomic events can be recognized as different by our apparatus is ultimately limited by quantum mechanical uncertainty. This dictates that any short time structure comes with a broad energy spectrum. In order that the streaks of two events separated by δt in time can be resolved they must be shifted with respect to each other by at least as much as their own spectral width $\delta W \approx \hbar/\delta t$. From this requirement we obtain the ATR resolving power as:

$$\delta t = \frac{T_0}{2\pi} \sqrt{\frac{\hbar\omega_L}{\Delta W_{\text{max}}}} \quad (2)$$

where ΔW_{max} stands for the energy shift suffered by the electron ejected at the peak of $A_L(t)$. Under our current experimental conditions ΔW_{max} can approach 25 eV before the onset of laser-induced ionization, yielding $\delta t < 100$ as for $T_0 = 2$ fs. The atomic transient recorder driven by red light and probing electrons with a kinetic energy near 100 eV is able to distinguish events within 100 as, constituting what is to our knowledge the shortest interval of time directly measurable to date.

Extension of the presented experiments to simultaneously exploring the temporal variation of emission intensity and momentum distribution of primary (photo) and secondary (Auger) electron emission will provide unprecedented insight into the excitation and relaxation dynamics of the electronic shell of atoms and molecules. The currently used time window of $T_0/2 \approx 1$ fs can be extended to several 10 fs by difference frequency generation with the few-cycle laser pulse while keeping the resolution of the ATR in the sub-fs regime, owing to $\delta t \propto \sqrt{T_0}$. With an extended time window a detailed characterization of trains of attosecond bursts (in contrast with the measurement of parameters averaged over the train¹⁹⁻²¹) will also become feasible. At near-keV excitation energies, the atomic transient recorder will allow time-domain metrology with a resolution approaching the atomic unit of time (24 as). □

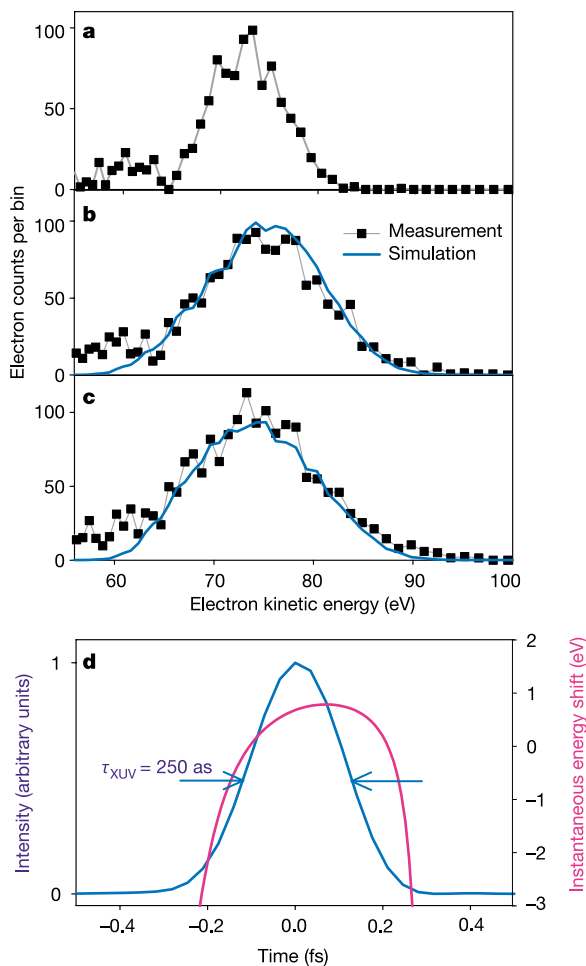


Figure 5 Selected streaked spectra from the ATR measurement of photoelectron emission from neon excited with a 93-eV sub-fs pulse (Fig. 4). **a**, Spectrum in the absence of laser field. **b**, **c**, Streaked spectra recorded at delays that imply coincidence of electron emission with (adjacent) zero transitions of $A_L(t)$, see arrows in Fig. 4. **d**, Temporal intensity profile and energy sweep of the sub-femtosecond XUV excitation pulse evaluated from the ATR measurements.

Methods

We sample electron emission from atoms using the experimental set-up that was described in detail¹ and employed recently for probing Auger electrons on a few-fs timescale²². The essential innovation here is that waveform-controlled few-cycle light now provides a reproducible excitation burst for accurate triggering and a reproducible streaking field for capturing sub-fs electron emission from atoms. For excitation, XUV bursts are produced from Ne atoms ionized by intense, few-cycle waveform-controlled light pulses⁵ in a process giving rise to high-order harmonics of the incident light for periodic (multi-cycle) pumping^{10,11}. The collinear XUV and laser beams are focused into a neon gas jet and delayed with respect to each other for the ATR measurements with a two-component Mo/Si broadband multilayer mirror¹ (radius of curvature = -70 mm) placed 2.5 m downstream from the source. The internal part of the laser beam (~3 mm in diameter) was blocked with a zirconium filter (transmitting the XUV pulse) in front of the Mo/Si mirror. The resultant annular laser beam (with a dark spot 3 mm in diameter in its centre) and the XUV beam (~3 mm in diameter) confined by the annular laser beam were focused by the external and internal sections of the two-component Mo/Si mirror¹, respectively. The two components can be aligned separately and translated with respect to each other along the optical axis of the mirror.

The reflectivity band of the multilayer extends from 85 V to 100 V with a peak reflectance of ~30% and a FWHM of ~9 V. Electrons ejected following the XUV excitation are collected within a narrow cone (<4°) aligned parallel to the laser and XUV polarization and analysed with a time-of-flight spectrometer. Two types of experiments have been implemented. First, by removing the zirconium filter we used the internal part of the Mo/Si mirror to focus both the XUV and the laser beam to eliminate any external source of fluctuations in the timing between the excitation and probing pulses (results summarized in Fig. 3). In the second type of studies, the XUV and the laser beam were reflected by different sections of the focusing mirror as described above. In these investigations the probing laser field could be delayed with respect to the XUV excitation pulse by the translation of the internal part of the mirror on a nanometre scale, yielding the data shown in Figs 4 and 5.

Received 24 October; accepted 5 December 2003; doi:10.1038/nature02277.

- Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
- Zewail, A. Femtochemistry: atomic-scale dynamics of the chemical bond (adapted from the Nobel Lecture). *J. Phys. Chem. A* **104**, 5660–5694 (2000).
- Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* **88**, 173903 (2002).
- Kitzler, M., Milosevic, N., Scrinzi, A., Krausz, F. & Brabec, T. Quantum theory of attosecond XUV pulse measurement by laser-dressed photoionization. *Phys. Rev. Lett.* **88**, 173904 (2002).
- Baltuska, A. *et al.* Attosecond control of electronic processes by intense light fields. *Nature* **421**, 611–615 (2003).
- Bradley, D. J., Liddy, B. & Sleat, W. E. Direct linear measurement of ultrashort light pulses with a picosecond streak camera. *Opt. Commun.* **2**, 391 (1971).
- Schelev, M. Ya., Richardson, M. C. & Alcock, A. J. Image-converter streak camera with picosecond resolution. *Appl. Phys. Lett.* **18**, 354 (1971).
- Kane, D. J. & Trebino, R. Characterization of arbitrary femtosecond pulses using frequency-resolved optical gating. *IEEE J. Quantum Electron.* **29**, 571–579 (1993).
- Sekikawa, T., Katsura, T., Miura, S. & Watanabe, S. Measurement of the intensity-dependent atomic dipole phase of a high harmonic by frequency-resolved optical gating. *Phys. Rev. Lett.* **88**, 193902 (2002).
- L'Huillier, A. & Balcou, Ph. High-order harmonic generation in rare gases with a 1-ps 1053-nm laser. *Phys. Rev. Lett.* **70**, 774–777 (1993).
- Macklin, J. J., Kmetec, J. D. & Gordon, C. L. III High-order harmonic generation using intense femtosecond pulses. *Phys. Rev. Lett.* **70**, 766–769 (1993).
- Schäfer, K. J., Yang, B., DiMauro, L. F. & Kulander, K. C. Above threshold ionization beyond the high harmonic cutoff. *Phys. Rev. Lett.* **70**, 1599–1602 (1993).
- Corkum, P. B. Plasma perspective on strong-field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Lewenstein, M., Balcou, Ph., Ivanov, M. Yu., L'Huillier, A. & Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).
- Christov, I. P., Murnane, M. M. & Kapteyn, H. C. High-harmonic generation of attosecond pulses in the “single-cycle” regime. *Phys. Rev. Lett.* **78**, 1251–1254 (1997).
- Kan, C., Burnett, N. H., Capjack, C. E. & Rankin, R. Coherent XUV generation from gases ionized by several cycle optical pulses. *Phys. Rev. Lett.* **79**, 2971–2974 (1997).
- de Bohan, A., Antoine, P., Milosevic, D. B. & Piraux, B. Phase-dependent harmonic emission with ultrashort laser pulses. *Phys. Rev. Lett.* **81**, 1837–1840 (1998).
- Tempea, G., Geissler, M. & Brabec, T. Phase sensitivity of high-order harmonic generation with few-cycle laser pulses. *J. Opt. Soc. Am. B* **16**, 669–674 (1999).
- Paul, P. M. *et al.* Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–1692 (2001).
- Mairesse, Y. *et al.* Attosecond synchronisation of high-harmonic soft X-rays. *Science* **302**, 1540–1543 (2003).
- Tzallas, P., Charalambidis, D., Papadogiannis, N. A., Witte, K. & Tsakiris, G. D. Direct observation of attosecond light bunching. *Nature* **426**, 267–271 (2003).
- Drescher, M. *et al.* Time-resolved atomic inner-shell spectroscopy. *Nature* **419**, 803–807 (2002).

Acknowledgements This work was sponsored by the Fonds zur Förderung der wissenschaftlichen Forschung (Austria), the Deutsche Forschungsgemeinschaft and the Volkswagenstiftung (Germany) and by the European Union's Human Potential Programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.K. (ferenc.krausz@tuwien.ac.at).

Giant magnetoresistance in organic spin-valves

Z. H. Xiong, Di Wu, Z. Vally Vardeny & Jing Shi

Department of Physics, University of Utah, Salt Lake City, Utah 84112, USA

A spin valve is a layered structure of magnetic and non-magnetic (spacer) materials whose electrical resistance depends on the spin state of electrons passing through the device and so can be controlled by an external magnetic field. The discoveries of giant magnetoresistance¹ and tunnelling magnetoresistance² in metallic spin valves have revolutionized applications such as magnetic recording and memory, and launched the new field of spin electronics³—‘spintronics’. Intense research efforts are now devoted to extending these spin-dependent effects to semiconductor materials. But while there have been noteworthy advances in spin injection and detection using inorganic semiconductors^{4–6}, spin-valve devices with semiconducting spacers have not yet been demonstrated. π -conjugated organic semiconductors may offer a promising alternative approach to semiconductor spintronics, by virtue of their relatively strong electron-phonon coupling⁷ and large spin coherence⁸. Here we report the injection, transport and detection of spin-polarized carriers using an organic semiconductor as the spacer layer in a spin-valve structure, yielding low-temperature giant magnetoresistance effects as large as 40 per cent.

π -conjugated organic semiconductors (OSEs) are a relatively new class of electronic materials that are revolutionizing important technological applications including information display⁹ (ref. 10 and references therein) and large-area electronics (ref. 11 and references therein), owing to their ability to be economically processed in large areas, their compatibility with low-temperature processing, the tunability of their electronic properties, and the simplicity of thin-film device fabrication. The virtually limitless flexibility of synthetic organic chemistry allows the fabrication of π -conjugated OSE structures with a degree of control unattainable with the conventional inorganic semiconductors. In addition, the OSEs have extremely weak spin-orbit interaction and weak hyperfine interaction, so that electron spin diffusion length is especially long⁸. These properties make them ideal for spin-polarized electron injection and transport applications.

Here we have chosen the small π -conjugated molecule 8-hydroxy-quinoline aluminium (Alq₃), most commonly used in organic light-emitting diodes (OLEDs)¹⁰, to serve as an OSE spacer in organic spin-valves, because it can easily be deposited as thin films and integrated with a variety of metallic electrodes. As shown in Fig. 1a, the vertical organic spin-valves that we fabricated consist of three layers: two ferromagnetic electrode films (FM₁ and FM₂, respectively) and the OSE spacer. By engineering the two FM electrodes to have different coercive fields (H_{c1} and H_{c2} , respectively), their magnetization directions can have either a parallel or anti-parallel alignment configuration when sweeping the external magnetic field, H ; this is essential for proving the spin-valve effect.

Intrigued by the possibility of spin-injection involving a π -conjugated OSE oligomer (sexi-thiophene, T₆)¹², where both the opposite FM electrodes were La_{0.67} Sr_{0.33} MnO₃ (LSMO), we have chosen LSMO for the bottom electrode (FM₁) and cobalt as the top electrode (FM₂) in our devices (Fig. 1a, b). LSMO is believed to be a half-metallic ferromagnet that possesses near-100% spin-polarization¹³. We note that unlike metallic FM films such as cobalt, nickel, iron or their alloys, the LSMO films are already stable against oxidation. In fact, our LSMO films have been cleaned and re-used

many times without any apparent degradation. After neat cleaning using acetone, the LSMO was introduced into the evaporation chamber with a base pressure of 5×10^{-7} torr, where the OSE film (Alq_3) was thermally evaporated. Without breaking the vacuum, we then deposited a thin (3–6 nm) cobalt (Co) film followed by an aluminium (Al) contact, using a shadow mask. The obtained active device area was about $2 \times 3 \text{ mm}^2$. We have fabricated several spin-valve devices with various OSE thicknesses d , between 130 to 250 nm. The OSE thickness was measured by an *in situ* thickness monitor and independently confirmed outside the chamber by scanning electron microscopy and thickness profilometry, respectively.

A schematic band diagram of a typical LSMO/ Alq_3 /Co device is shown in Fig. 1c. In the rigid band approximation—that is, without taking into account the relaxation and polarization energy associated with charge injection—the highest occupied molecular orbital (HOMO) of Alq_3 lies about 0.9 eV below the Fermi levels, E_F , of the FM electrodes, whereas the lowest unoccupied molecular orbital (LUMO) lies about 2.00 eV above E_F (refs 14, 15). At low applied bias voltages V , holes are injected from the anode into the HOMO level of the OSE mainly by tunnelling through the bottom potential barrier. In addition, the similar work function value ϕ of the two electrodes (Fig. 1c) leads to a symmetric current–voltage I – V response (Fig. 1d). For fabricated devices with $d > 100 \text{ nm}$ we found that the I – V characteristic was nonlinear with a weak temperature dependence (Fig. 1d), indicative of carrier injection by tunnelling. Control devices with similar OSE thickness, in which ITO replaced the LSMO bottom electrode, showed electroluminescence and also a conductivity detected magnetic resonance at $g \approx 2$ (see Supplementary Information), indicating carrier injection into the OSE. In addition, at low V we measured a typical resistance of

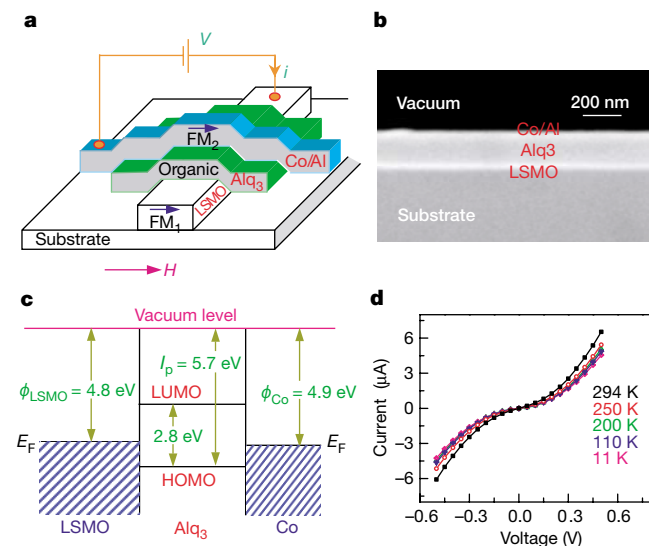


Figure 1 The structure and transport properties of the fabricated organic spin-valve devices. **a**, Schematic representation of a typical device that consists of two FM electrodes (FM_1 and FM_2) and an OSE spacer. Spin-polarized electrical current I flows from FM_1 (LSMO), through the OSE spacer (Alq_3), to FM_2 (Co) when a positive bias V is applied. An in-plane magnetic field, H , is swept to switch the magnetization directions of the two FM electrodes separately. **b**, Scanning electron micrograph of a functional organic spin-valve consisting of a 60-monolayer-thick LSMO film, a 160-nm-thick Alq_3 spacer, a 3.5-nm-thick Co electrode and a 35-nm-thick Al electrode. **c**, Schematic band diagram of the OSE device in the rigid band approximation showing the Fermi levels and the work functions of the two FM electrodes, LSMO and Co, respectively, and the HOMO–LUMO levels of Alq_3 . **d**, I – V response of the organic spin-valve device with $d = 200 \text{ nm}$ at several temperatures.

10^4 – $10^5 \Omega$ that depends on the deposition rate and thickness of the Co electrode; such resistance is also consistent with a dominant pinhole-free organic spacer. Devices with $d < 100 \text{ nm}$, however, showed a linear I – V response and lack of both conductivity detected magnetic resonance and electroluminescence, leading us to believe that these devices have an ‘ill-defined’ layer of up to 100 nm that may contain pinholes and Co inclusions. These findings suggest that the OSE spacers in the spin-valve devices fabricated with $d > 100 \text{ nm}$ may be composed of two sublayers: one sublayer with a thickness d_0 of the order of 100 nm immediately below the Co electrode that contains Co inclusions owing to the interdiffusion; and a second sublayer of neatly deposited Alq_3 between this defected sublayer and the LSMO film, having a thickness $d - d_0$, in which carrier transport is dominated by carrier drift.

The magnetoresistance of the fabricated devices was measured in a closed-cycle refrigerator from 11 to 300 K by sending a current through the two interfaces, while varying an external in-plane magnetic field, H (Fig. 1a). The hysteresis loops of the magnetization versus H for the FM electrodes were measured by

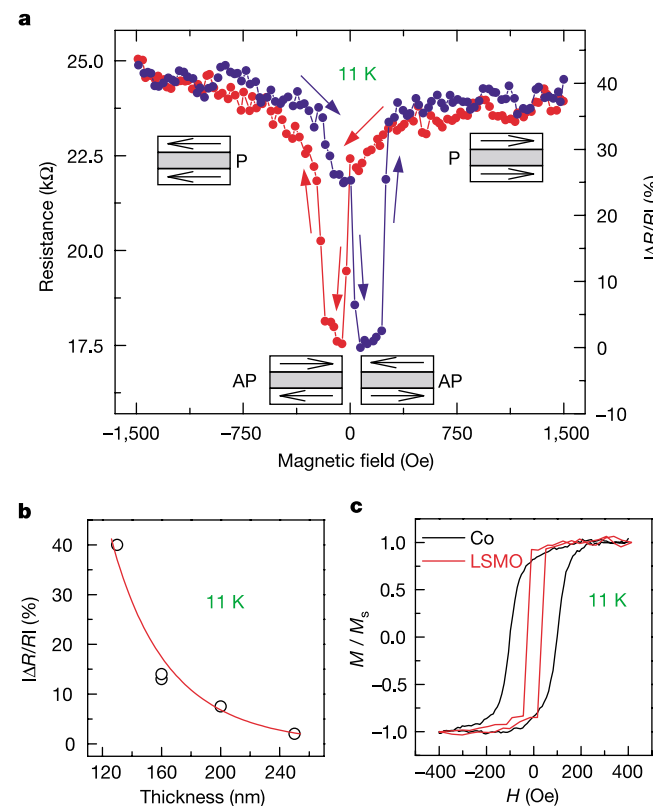


Figure 2 The magneto-transport response of the OSE spin-valve devices. **a**, GMR loop of a LSMO (100 nm)/ Alq_3 (130 nm)/Co (3.5 nm) spin-valve device measured at 11 K. The blue (red) curve denotes GMR measurements made while increasing (decreasing) H . The anti-parallel (AP) and parallel (P) configurations of the FM magnetization orientations are shown in the insets at low and high H , respectively. The electrical resistance of the device is higher when the magnetization directions in FM_1 and FM_2 films are parallel to each other. **b**, The GMR value of a series of LSMO/ Alq_3 /Co devices with different d . The line fit through the data points was obtained using the spin diffusion model, equation (1), with three adjustable parameters, as explained in the text. All devices were fabricated on the same LSMO film. **c**, Magnetic hysteresis loops measured using MOKE for the LSMO electrode ($H_{c1} \approx 30 \text{ Oe}$), and Co film deposited on Alq_3 under the same conditions as that for the Co electrode in the actual spin-valve devices ($H_{c2} \approx 150 \text{ Oe}$). These coercive fields are seen in **a** when the resistance switches from parallel to the anti-parallel configuration, and vice versa.

the magneto-optic Kerr effect (MOKE) over the same temperature range. For comparison, we also measured separately the magnetoresistance (ΔR) of each FM electrode, as well as the OSE spacer on a variety of fabricated control devices: in all cases we found ΔR to be extremely small (less than 0.1% of the ΔR in the spin-valve devices) and devoid of any hysteresis with H (see Supplementary Information).

Figure 2a shows a typical magnetoresistance loop obtained in an LSMO/Alq₃/Co spin-valve device with $d = 130$ nm; the magnetoresistance curves of three other devices having larger d are shown in the Supplementary Information and their $\Delta R/R$ is summarized in Fig. 2b. A sizeable $\Delta R/R$ of 40%, which is a giant magnetoresistance (GMR) response comparable to that obtained in metallic GMR spin-valves^{1,16}, is observed at 11 K. The GMR of the devices with larger d is progressively smaller, but still measurable up to $d = 250$ nm (Fig. 2b). MOKE measurements performed on the LSMO bottom electrode of the device in Fig. 1a indicate that the coercive field of the LSMO film is $H_{c1} \approx 30$ Oe at 11 K (Fig. 2c). Although the top Co electrode of this device is not accessible to MOKE owing to the Al top contact, the coercive field of a Co film of the same thickness deposited on Alq₃ under similar conditions was nevertheless measured to be $H_{c2} \approx 150$ Oe at 11 K, much greater than that of the LSMO. Clearly then, the magnetization orientations in the two FM electrodes are anti-parallel to each other when the external field H is between H_{c1} and H_{c2} ; in contrast, the magnetization orientations are parallel to each other when the field strength

$H > H_{c2}$ (Fig. 2a insets). Therefore, the observed GMR hysteresis is undoubtedly due to the spin-valve effect. We note that the resistance in the anti-parallel alignment is lower than that in the parallel alignment, which is opposite to the spin-valve effect usually obtained using two identical FM electrodes, or two different 'd-band' metallic FM electrodes in some devices¹⁶. This 'inverse magnetoresistance' was also previously seen in LSMO/SrTiO₃/Co (ref. 17) and LSMO/Ce_{0.69}La_{0.31}O_{1.845}/Co (ref. 18) magnetic tunnel junction devices having extremely thin insulating spacers (~ 2 nm). The inverse magnetoresistance is believed to originate from the negative spin polarization of the Co d -band, in which the density-of-states of the majority-spin sub-band at the Fermi level is smaller than that of the minority-spin sub-band.

We may analyse the obtained GMR effect and its dependence on d using a simple injection and diffusion model. In conventional magnetic tunnel junction devices with a very thin insulating tunnel barrier, the Jullière model¹⁹ has often been used to analyse the tunnelling magnetoresistance. In the present organic spin-valves, the neatly deposited OSE sublayer with thickness $d-d_0$ is actually so thick (>30 nm) that simple quantum mechanical tunnelling through it is not a viable possibility. Although the detailed physical picture is lacking, we nevertheless assume that there exists a potential barrier for spin injection at the Co/OSE interface (Fig. 1c), which may be self-adjusted⁷. Once carriers are injected through this interface they easily reach the neat sublayer, where they drift under the influence of the electric field towards the other

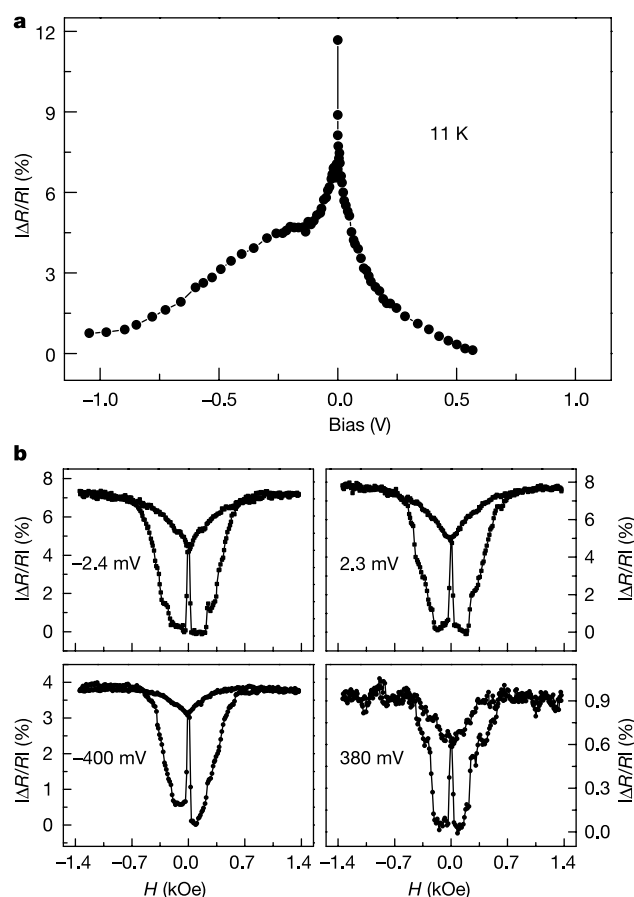


Figure 3 Bias voltage dependence. **a**, $|\Delta R/R|$ for a LSMO/Alq₃/Co device with $d = 160$ nm measured at 11 K, as a function of V . **b**, Spin-valve related GMR loops at four different V values.

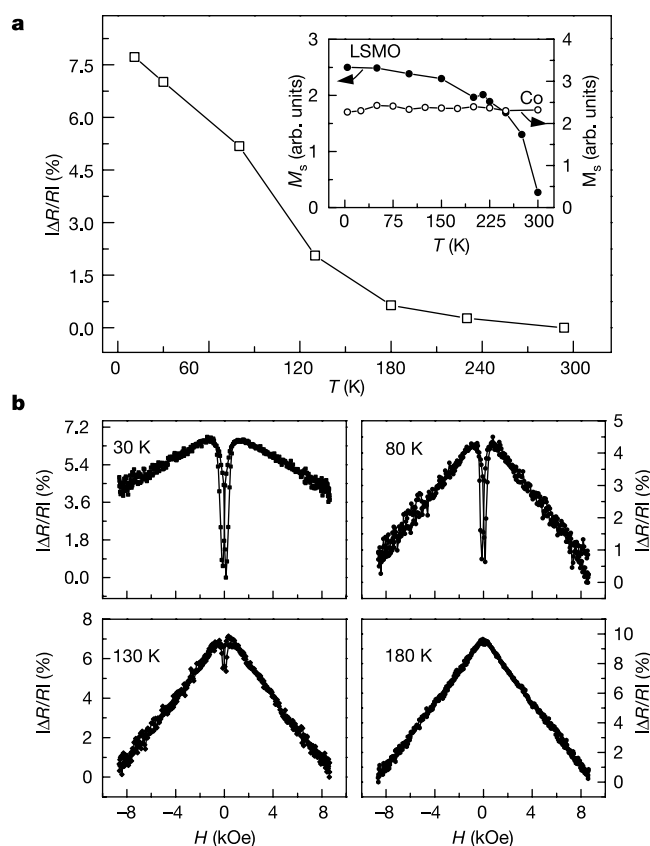


Figure 4 Temperature dependence. **a**, $|\Delta R/R|$ measured at $V = 2.5$ mV as a function of temperature for the same device as in Fig. 3. The inset shows the magnetizations of the Co and LSMO electrodes versus T , measured using MOKE. **b**, The low-field GMR hysteresis loop, and the high-field magnetoresistance measured at various temperatures as indicated. M_s , saturation magnetization.

interface, from which they can be extracted. As the injected carriers reach the end of the ill-defined sublayer, the spin polarization is p_1 ; it further decays in the remaining neatly deposited sublayer with a surviving probability $\exp[-(z-d_0)/\lambda_s]$, where z is the drift/diffusion distance along the normal direction to the interface, and λ_s is the spin diffusion length in the neatly deposited OSE sublayer. The spin polarization p is defined as $p = \frac{N_{\uparrow} - N_{\downarrow}}{N_{\uparrow} + N_{\downarrow}}$, where $N_{\uparrow}$ ($N_{\downarrow}$) is the carrier density in the majority (minority) spin state. We express the thickness dependence of the GMR magnitude, $\Delta R/R$, which is the maximum relative change in electrical resistance R within the spin-valve hysteresis loop, assuming no loss of spin memory at the interfaces owing to the self-adjusting capability of the OSE⁷:

$$\frac{\Delta R}{R} = \frac{R_{AP} - R_P}{R_{AP}} = \frac{2p_1 p_2 e^{-(d-d_0)/\lambda_s}}{1 + p_1 p_2 e^{-(d-d_0)/\lambda_s}} \quad (1)$$

where R_{AP} and R_P denote R in the anti-parallel and parallel magnetization configurations, respectively. For inverse magnetoresistance, $R_{AP} < R_P$; therefore $\Delta R/R$ is negative, according to equation (1). We used equation (1) with three adjustable parameters, namely $p_1 p_2$, d_0 and λ_s to fit the data for $\Delta R/R$ in Fig. 2b. The good agreement shown in Fig. 2b was obtained with the following parameters: $p_1 p_2 = -0.32$; $d_0 = 87$ nm, which is close to the lower limit of d below which the I - V response becomes linear; and $\lambda_s = 45$ nm, which is a reasonable value for λ_s in the neatly deposited OSE sublayer. The obtained λ_s is smaller than what was extracted for T_6 (ref. 12), possibly owing to the aluminium element in Alq_3 that may increase the spin-orbit coupling in this OSE. However, λ_s is similar to what was extracted from spin-valve devices of single-walled carbon nanotubes²⁰.

We measured the dependences of $\Delta R/R$ of the organic spin-valve devices on the bias voltage V and temperature T . A series of GMR hysteresis curves for different V are displayed in Fig. 3b for the $d = 160$ nm device; similar dependence was obtained in devices having other OSE thicknesses. $\Delta R/R$ strongly depends on V , as summarized in Fig. 3a; it peaks at $V \approx 0$ and decreases as the applied voltage increases. In addition, the V dependence exhibits a clear asymmetry; it decays more slowly with V for negative voltages, which is similar to what was observed in magnetic tunnel junctions with LSMO and Co or CoFe electrodes^{17,21}.

$\Delta R/R$ also strongly depends on T . A series of GMR hysteresis is shown in Fig. 4b; the complete temperature dependence of $\Delta R/R$ is summarized in Fig. 4a. From equation (1) the decrease of $\Delta R/R$ at high temperatures may contain two contributions. One contribution is the decreased spin polarization of the LSMO injecting electrode, p_1 at high T and the other is the decreased λ_s of the neatly deposited OSE sublayer⁸. A decrease in p_1 is expected for LSMO owing to its low Curie temperature of about 300 K (Fig. 4a, inset). The temperature dependence of $\Delta R/R$, however, does not qualitatively follow that of the magnetization of the LSMO itself; $\Delta R/R$ vanishes below the Curie temperature of the LSMO. This indicates that λ_s in Alq_3 has a stronger temperature dependence than that of the LSMO magnetization. To check this conclusion, we also fabricated a spin-valve device in which the LSMO bottom electrode was replaced by a Fe electrode having a much higher Curie temperature than LSMO. Although both FM electrodes of this device maintain their ferromagnetic properties well above room temperature, we still observed a decrease of the GMR with T , similar to that displayed in Fig. 4a. In addition, we also measured the T dependence of the spin relaxation time in Alq_3 using the spin-1/2 photoluminescence detected magnetic resonance at $g \approx 2$ as a function of temperature. We found that the spin-1/2 resonance signal decreases with T in much the same way as the decrease with T of the spin-valve GMR in our devices (Fig. 4a). This shows first, that spin-1/2 carriers in the OSE are responsible for the GMR response in our devices; and, second, that the decreased GMR at

high temperatures is due to the increase of the spin relaxation rate in the OSE.

At increased T we found another interesting effect; an increase in high-field magnetoresistance (HFMR) accompanies the decrease of the low-field spin-valve related GMR (Fig. 4b). We have also observed a similar but much smaller HFMR in the LSMO film itself (Supplementary Information). However, the spin-valve device resistance is several orders of magnitude greater than that of the LSMO electrode. Therefore, the apparent HFMR cannot simply be explained by the change in the serial resistance of the LSMO electrode. Similar HFMR response was also reported in some other devices using LSMO electrodes¹⁸; and it can also explain the recently measured room-temperature resistance difference between two magnetic fields in planar LSMO/ T_6 /LSMO devices¹². In the latter devices, the relative magnetization orientations in the two identical electrodes should not depend on H ; therefore the observed magnetoresistance may be of the same origin as that of this HFMR, but is not due to the conventional spin-valve effect.

This study shows that spin-polarized carrier injection, transport and detection, which are the main ingredients of spintronics, can be successfully achieved using π -conjugated OSE. This may initiate a variety of exciting new applications in organic spintronics such as spin-OLEDs, enabled by the functionalities of the OSE. $\square$

Received 13 November; accepted 23 December 2003; doi:10.1038/nature02325.

1. Baibich, M. N. *et al.* Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
2. Moodera, J., Kinder, L., Wong, T. & Meservey, R. Magnetic tunnel junction. *Phys. Rev. Lett.* **74**, 3273–3276 (1995).
3. Wolf, S. A. *et al.* Spintronics: a spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
4. Kikkawa, J. M. & Awschalom, D. D. Lateral drag of spin coherence in gallium arsenide. *Nature* **397**, 139–141 (1999).
5. Ohno, Y. *et al.* Electrical spin injection in a ferromagnetic semiconductor heterostructure. *Nature* **420**, 790–792 (1999).
6. Hanbicki, A. T. *et al.* Efficient electrical spin injection from a magnetic metal/tunnel barrier contact into a semiconductor. *Appl. Phys. Lett.* **80**, 1240–1242 (2002).
7. Xie, S. J., Ahn, K. H., Smith, D. L., Bishop, A. R. & Saxena, A. Ground-state properties of ferromagnetic metal/conjugated polymer interfaces. *Phys. Rev. B* **67**, 125202 (2003).
8. Krinichnyi, V. I. 2-mm waveband electron paramagnetic resonance spectroscopy of conducting polymers. *Synth. Met.* **108**, 173–222 (2000).
9. Friend, R. H. *et al.* Electroluminescence in conjugated polymers. *Nature* **397**, 121–128 (1999).
10. Forrest, S., Burrows, P. & Thompson, M. The dawn of organic electronics. *IEEE Spectr.* **37**, 29–34 (2000).
11. Voss, D. Cheap and cheerful circuits. *Nature* **407**, 442–444 (2000).
12. Dediu, V., Murgia, M., Maticotta, F. C., Taliani, C. & Barbanera, S. Room temperature spin polarised injection in organic semiconductor. *Solid State Commun.* **122**, 181–184 (2002).
13. Bowen, M. *et al.* Nearly total spin polarization in $\text{La}_{2/3}\text{Sr}_{1/3}\text{MnO}_3$ from tunneling experiments. *Appl. Phys. Lett.* **82**, 233–235 (2003).
14. Lee, S. T., Hou, X. Y., Mason, M. G. & Tang, C. W. Energy level alignment at Alq_3 /metal interfaces. *Appl. Phys. Lett.* **72**, 1593–1595 (1998).
15. Arisi, E. *et al.* Organic light emitting diodes with spin polarized electrodes. *J. Appl. Phys.* **93**, 7682–7683 (2003).
16. Diény, B. *et al.* Giant magnetoresistance of magnetically soft sandwiches: dependence on temperature and on layer thicknesses. *Phys. Rev. B* **45**, 806–813 (1992).
17. De Teresa, J. M. *et al.* Inverse tunnel magnetoresistance in $\text{Co/SrTiO}_3/\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$: new ideas on spin-polarised tunnelling. *Phys. Rev. Lett.* **82**, 4288–4291 (1999).
18. De Teresa, J. M. *et al.* Role of metal-oxide interface in determining the spin polarization of magnetic tunnel junctions. *Science* **286**, 507–509 (1999).
19. Jullière, M. Tunneling between ferromagnetic films. *Phys. Lett. A* **54**, 225–226 (1975).
20. Tsukagoshi, K., Alphenaar, B. W. & Ago, H. Coherent transport of electron spin in a ferromagnetically contacted carbon nanotube. *Nature* **401**, 572–574 (1999).
21. Hayakawa, J. *et al.* The origin of bias-dependence in $\text{CoFe/SrTiO}_3/\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ magnetic tunnel junctions. *J. Appl. Phys.* **91**, 8792–8794 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank X.G. Li at the University of Science and Technology of China for providing the LSMO films and C. Taliani, V. Dediu, V. Burtman and D. Smith for discussions. We also thank F. J. Wang, C. G. Yang, C. Z. Liu and M. DeLong for assistance with the measurements. This work was supported in part by the National Science Foundation, Petroleum Research Foundation, DARPA, and the Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (jshi@physics.utah.edu).

Bottom water warming in the North Pacific Ocean

Masao Fukasawa¹, Howard Freeland², Ron Perkin²,
Tomowo Watanabe^{3*}, Hiroshi Uchida¹ & Ayako Nishina⁴

¹Ocean Observation and Research Department, Japan Marine Science and Technology Centre, Yokosuka, 237-0061, Japan

²The Institute of Ocean Sciences, Sidney, British Columbia, V8L 4B2, Canada

³Far-fisheries Laboratory, Japan Fisheries Agency, Shimizu, 424-8633, Japan

⁴Faculty of Fisheries, Kagoshima University, Kagoshima, 890-0056, Japan

* Present address: Central Fisheries Laboratory, Japan Fisheries Agency, Japan

Observations of changes in the properties of ocean waters have been restricted to surface¹ or intermediate-depth waters^{2,3}, because the detection of change in bottom water is extremely difficult owing to the small magnitude of the expected signals. Nevertheless, temporal changes in the properties of such deep waters across an ocean basin are of particular interest, as they can be used to constrain the transport of water at the bottom of the ocean and to detect changes in the global thermohaline circulation. Here we present a comparison of a trans-Pacific survey completed in 1985 (refs 4, 5) and its repetition in 1999 (ref. 6). We find that the deepest waters of the North Pacific Ocean have warmed significantly across the entire width of the ocean basin. Our observations imply that changes in water properties are now detectable in water masses that have long been insulated from heat exchange with the atmosphere.

Between 1991 and 1996 more than 100 hydrographic lines were surveyed to the highest standards then available. One such line scheduled for survey was designated World Ocean Circulation Experiment (WOCE) P1, across the subarctic Pacific. This line was not completed within the observation period for the WHP (WOCE Hydrographic Program) and had actually been completed in 1985. A revisit of P1 was scheduled as part of the Japanese Sub-Arctic Gyre Experiment and actually completed collaboratively between several Japanese agencies and the Institute of Ocean Sciences in Canada. Here we present results suggesting changes in the properties of the densest water in the North Pacific, which corresponds to the modified NADW (North Atlantic Deep Water) θ class water defined in the Samoa passage⁷.

The WHP line P01, located mainly along 47° N (see Fig. 1) from coast to coast, was surveyed in 1985 (refs 4, 5) and will be referred to

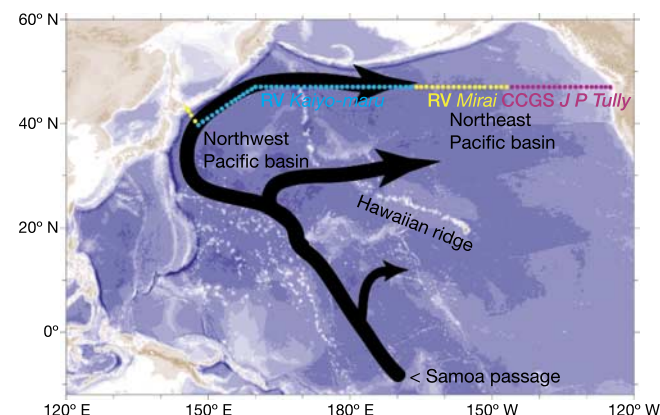


Figure 1 The location of stations occupied for WHP Line P1 revisit in 1999. The stations are colour-coded according to the ship completing that part of the survey. The dark arrows are adapted from ref. 17 to illustrate the dominant pathways by which bottom water from the Samoa passage is carried into the northeast Pacific basin.

here as P1_85. The P1 revisit (P1_99) was scheduled to examine the long-term variability in the subarctic Pacific⁶. The revisit survey was composed of four separate cruises. The accuracy of temperature and salinity measurement for the 1985 cruise is reported as 0.001 °C and 0.002 p.s.u. respectively⁴. Across the entire section in 1999, the accuracy of temperature (see Methods) and salinity measurement is better than 0.001 °C and about 0.001 p.s.u., respectively. We therefore consider differences smaller than 0.002 °C and 0.003 p.s.u. as being insignificant. These estimates are conservative and are thus more likely to discard real differences than to accept insignificant differences.

Figure 2a and b are cross-sections of the differences in temperature (ΔT) and salinity (ΔS) computed as 1999 conditions minus 1985 conditions. Regions of insignificant difference are not coloured. Below 3,000 m there is no area where the salinity differences are significant. Ignoring the shallower levels where differences due to seasonal variability are large, we note the following observations: (1) we see no obvious mismatches where the different sections of P1_99 are joined, suggesting that the difference significance levels truly are conservative. (2) Between 500 and 3,000 m there are large signals in both ΔT and ΔS that are negatively correlated with each other. (3) Below 5,000 m and extending to

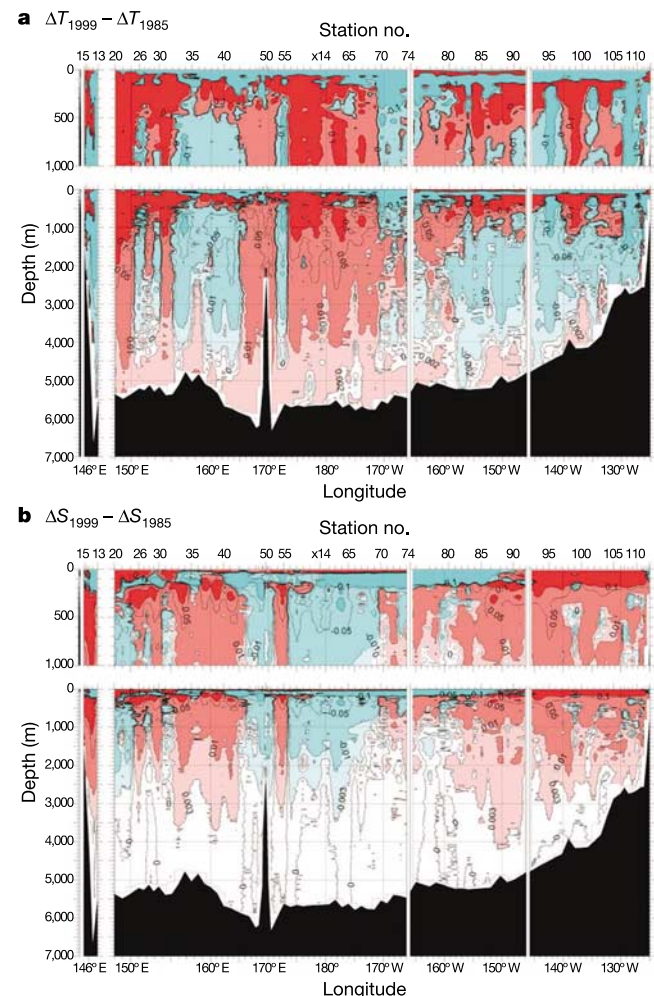


Figure 2 The difference in potential temperature (a) and salinity (b) between the two surveys of line P1 in 1999 and 1985, plotted against depth. The temperature and salinity differences are in °C and practical salinity units (p.s.u.), respectively. The contour intervals are variable and chosen to illuminate changes in the deeper waters. Dark red, very significant increases; pale red, less significant increases. Dark blue, very significant decreases; pale blue, less significant decreases.

the bottom there is a basin-wide warming of about 0.005°C with no significant change in the salinity field.

Figure 3 shows ΔS again but plotted against longitude and the density variable σ_4 from values of 45.5 to the largest value observed in the section. The value $\sigma_4 = 45.5$ occurs at an average depth of about 1,400 m. We note that ΔS is larger in the Canadian part of this survey (from the CCGS *John P. Tully*), which might possibly be indicative of a lower quality of salinity determinations. However, nowhere in the section are the values of ΔS greater than our suggested detection limit of 0.003 p.s.u., further suggesting that our proposed detection limit is conservative. In the mid-depth region the values of ΔS are small compared with the values observed on pressure surfaces, suggesting that temperature and salinity changes are largely density-compensating. Thus these changes can largely be effected by a redistribution of water masses and do not require systematic changes in the properties of such water masses to be greater than 0.002°C or 0.005 p.s.u..

In contrast, the changes below 5,000 m do not show the longitudinally banded structure that characterizes the mid-depth changes, and in particular show significant changes in temperature but not salinity. Thus these changes cannot be effected by vertical mass redistributions. Though the deep-water temperature changes are evident in all regions of the section the changes are particularly strong west of 160°W and especially near the Emperor seamount chain near 168°E . The deep and bottom waters of the North Pacific are filled with North Pacific Deep Water and modified North Atlantic Deep Water (mNADW), which composes the upper portion of the Lower Circumpolar Deep Water (LCDW). The NPDW and mNADW masses are characterized by potential temperatures lower than 2°C and 1.2°C , respectively, in the Samoan passage.

Figure 4 is constructed to clarify the changes taking place in the bottom water. For a series of potential temperature classes we compute a series of areas $A(\theta_i)$ defined as the total area along the section between $\theta_i - 0.005^{\circ}\text{C}$ and $\theta_i + 0.005^{\circ}\text{C}$, then look at the difference field $\Delta A(\theta_i) = A_{1999}(\theta_i) - A_{1985}(\theta_i)$. The function $\Delta A(\theta_i)$ is plotted in the left panel of Fig. 4 and in the right panel we plot $I(\theta_k)$, where:

$$I(\theta_k) = \sum_{i=1}^k \Delta A(\theta_i)$$

and $i = 1$ denotes the coldest temperature bin. This clearly shows that the colder portion of the mNADW (that is, $\theta < 1.1^{\circ}\text{C}$) decreased its presence along 47°N between 1985 and 1999, whereas

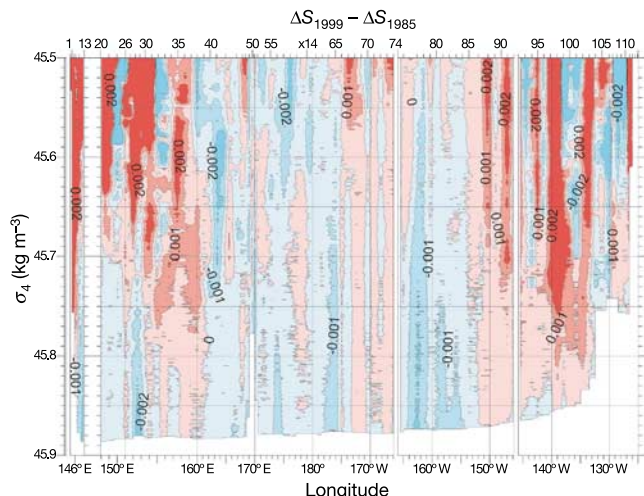


Figure 3 The difference in salinity between 1999 and 1985 for line P1, as in Fig. 2b, but plotted with the density parameter σ_4 as the vertical axis. σ_4 is a density anomaly referenced to a standard pressure of 4,000 decibar.

the warmer portion defined as $1.2^{\circ}\text{C} \geq \theta \geq 1.1^{\circ}\text{C}$ increased its presence. The integral from the bottom vanishes at 1.2°C , showing that the total volume of mNADW has not changed. Throughout the NPDW mass the function $\Delta A(\theta_i)$ is close to zero, indicating no significant changes.

If we assume that $4 \times 10^8 \text{ m}^2$ of the lower-area water (mean temperature 1.05°C) is replaced by the upper-area water at a temperature of 1.15°C , we can estimate a heat input of about $1.5 \times 10^{14} \text{ J}$ per metre width of the section. This is not a gigantic amount of heat. The temperature increase could be accommodated by increasing the heat input through the ocean bottom by about 0.080 W m^{-2} west of 160°W . Sclater and Parsons⁸ review the mechanism of heat flow through the ocean bottom and show that the anticipated mean heat flow varies with the age of the crust, with the largest values being associated with very young crust near spreading ridges. Most of the WHP Section P1 is over very old oceanic crust with a mean heat flux of 0.042 W m^{-2} . This is close to estimates from hydrographic data⁹ of about 0.05 W m^{-2} . So, to accommodate the observed changes using geothermal heat input would require an increase above normal of almost 200% of the expected value. This is unlikely to have occurred and so we conclude that a large part of the warming must be explained by advective changes.

At depths where the warming was observed, vertical variations in salinity stratification are too small to detect any significant signal of a change in bottom water, however, dissolved oxygen shows a sharper stratification than salinity, providing an alternative view. Dissolved oxygen values analysed for the warmer portion ($1.2^{\circ}\text{C} \geq \theta \geq 1.1^{\circ}\text{C}$) and those for the colder portion ($1.1^{\circ}\text{C} \geq \theta$) were averaged zonally. Averaged dissolved oxygen values were determined to be 148 and $147 \mu\text{mol kg}^{-1}$ for the warmer portion in P1_85 and P1_99, respectively. For the colder portion, we observe 158 and $157 \mu\text{mol kg}^{-1}$ in P1_85 and P1_99, respectively. Thus, no significant change in the relationship between potential temperature and dissolved oxygen has occurred for either the warmer or colder portions of the mNADW. Again, it is unlikely that the observed warming can be explained by geothermal heating. We conclude therefore that the changes observed result from changes in the oceanographic circulation.

We have shown that the water in the densest water masses warmed along 47°N during the period 1985–1999 and conclude that a change in the oceanography is needed to account for this evolution. The warming was brought about by a decrease in the volume of the lower part of the mNADW at least along 47°N . The distribution of dissolved oxygen in the deep layers of the North

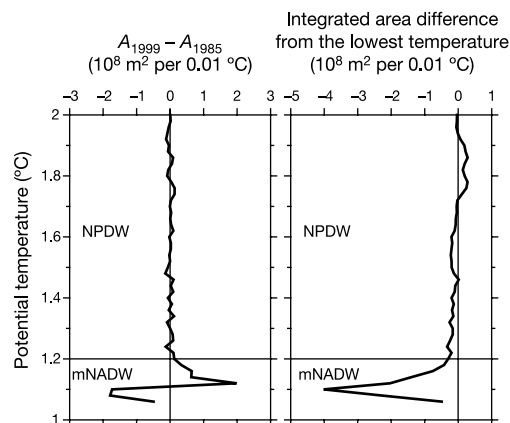


Figure 4 The difference in area between the 1999 and 1985 surveys in the areas along WHP Line P1 occupied by different potential temperatures. The bin size for potential temperature, used to define each area, is 0.01°C . See text for full description of calculations and equations used.

Pacific¹⁰ suggests that the bottom water in the western North Pacific is fed from the Samoa passage, spreading northwards through the Wake Island passage (see Fig. 1). In the South Pacific, the source region of the bottom water in the North Pacific, warming of the NADW has previously been suggested^{7,11}, all reporting changes in water masses in the Samoa passage, within the relatively recent time period when data are available.

These suggest that there is significant variability in the supply of Antarctic Bottom Water. Another recent study¹² concluded that mid-depth Southern Ocean temperatures have warmed by as much as 0.17 °C since the 1950s, when the modern accumulation of hydrographic data in this area started, within the Antarctic Circumpolar Current. This warming, together with the warming trend of the global ocean¹ is naturally expected to affect the Antarctic overturning system. It is interesting to speculate whether or not the changes we observe are directly related to the changes being observed in the South Pacific, and this will presumably be answered only after further revisits of WHP lines in the Pacific Ocean. □

Methods

The 1999 surveys all used SeaBird 911 CTD systems and field calibrations were completed using Guildline salinometers. The temperature sensors on the Japanese systems were calibrated within a month before and after each cruise and the accuracy of the temperature measurement is known to be better than 0.0007 °C. The temperature sensor on the Canadian system was calibrated three weeks before the survey to 0.0006 °C. Salinity observations were corrected using the standard seawater batch correction. This involved corrections of −0.0014 and −0.0015 to the Japanese observations, and −0.0009 to the Canadian data. We note that the separate segments of the survey did not match in the deep water until these corrections were made. Before differences were computed between the 1985 and 1999 surveys we converted the 1985 temperature observations from the reported values in the IPTS68 temperature scale to ITS90 (ref. 13), and applied the standard seawater batch correction^{14–16}, correcting salinities by +0.0012.

Received 22 July; accepted 31 December 2003; doi:10.1038/nature02337.

- Levitus, S., Antonov, J. L., Boyer, T. P. & Stephens, C. Warming of the world ocean. *Science* **287**, 2225–2229 (2000).
- Wong, A. P., Bindoff, N. L. & Church, J. A. Large-scale freshening of intermediate waters in the Pacific and Indian Oceans. *Nature* **400**, 440–443 (1999).
- Roemmich, D. & Wunsch, C. Apparent changes in the climatic state of the deep North Atlantic Ocean. *Nature* **307**, 447–450 (1984).
- Talley, L. D., Martin, M., Salameh, P. & the Oceanographic Data Facility. *Trans-Pacific Section in the Subpolar Gyre (TPS47): Physical, Chemical, and CTD Data, R/V Thomas Thompson TT190, 4 August 1985–7 September 1985* 1–246 (Ref. 88–9, Scripps Inst. Of Oceanography, La Jolla, CA, 1988).
- Talley, L. D., Joyce, T. & deSzoeke, R. A. Trans-Pacific sections at 47°N and 152°W: distribution of properties. *Deep-Sea Res. (Suppl.)* **38**, S63–S82 (1991).
- Uchida, H., Fukasawa, M. & Freeland, H. J. *WHP P01 Revisit Data Book 1–73* (Jamstec, Yokosuka, 2002).
- Johnson, G. C., Rudnick, D. L. & Taft, B. A. Bottom water variability in the Samoa Passage. *J. Mar. Res.* **52**, 177–195 (1994).
- Sclater, J. G. & Parsons, B. Oceans and continents: similarities and differences in the mechanisms of heat loss. *J. Geophys. Res.* **85**, 11535–11552 (1981).
- Joyce, T. M., Warren, B. A. & Talley, L. D. The geothermal heating of the abyssal subarctic Pacific Ocean. *Deep-Sea Res.* **33**, 1003–1005 (1986).
- Mantyla, A. & Reid, J. L. Abyssal characteristics of the world ocean waters. *Deep-Sea Res.* **30**, 805–833 (1983).
- Johnson, G. C. & Orsi, A. H. Southwest Pacific water-mass exchanges between 1968/69 and 1990/91. *J. Clim.* **10**, 306–316 (1997).
- Gille, S. H. Warming of the southern ocean since the 1950s. *Science* **295**, 1275–1277 (2002).
- Preston-Thomas, H. The international temperature scale of 1990. *Metrologia* **27**, 3–10 (1990).
- Mantyla, A. W. Standard seawater comparison updated. *J. Phys. Oceanogr.* **17**, 543–548 (1987).
- Aoyama, M., Joyce, T., Kawano, T. & Takatsuki, Y. Standard seawater comparison up to P129. *Deep-Sea Res.* **49**, 1103–1114 (2002).
- Kawano, T., Takatsuki, Y., Imai, J. & Aoyama, M. Seawater and quality evaluation of the standard seawater supplied in a bottle. [In Japanese with English abstract]. *J. Jpn Soc. Mar. Surv. Tech.* **13**, 11–18 (2001).
- Reid, J. L. On the total geostrophic circulation of the Pacific Ocean: flow patterns, tracers and transports. *Prog. Oceanogr.* **39**, 263–352 (1997).

Acknowledgements We thank the officers and crew of the three research vessels (RV *Kaiyo-maru* of JFA, RV *Mirai* of JAMSTEC, and CCGS *John P. Tully*) and all technical assistants. The Canadian contribution was funded by the Strategic Science Fund of the Department of Fisheries and Oceans, and the Japanese contributions were funded by the Promotional Foundation for Science and Technology of the Science and Technology Agency of Japan (now known as the Ministry of Education, Culture, Sports, Science and Technology).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F. (fks@jamstec.go.jp) or H.F. (FreelandHj@pac.dfo-mpo.gc.ca).

Rejuvenation of the lithosphere by the Hawaiian plume

Xueqing Li¹, Rainer Kind^{1,2}, Xiaohui Yuan¹, Ingo Wölbern¹ & Winfried Hanka¹

¹GeoForschungsZentrum Potsdam, Telegrafenberg, 14473 Potsdam, Germany

²Freie Universität Berlin, FR Geophysik, Malteserstrasse 74–100, 12249 Berlin, Germany

The volcanism responsible for creating the chain of the Hawaiian islands and seamounts is believed to mark the passage of the oceanic lithosphere over a mantle plume^{1,2}. In this picture hot material rises from great depth within a fixed narrow conduit to the surface, penetrating the moving lithosphere³. Although a number of models describe possible plume–lithosphere interactions⁴, seismic imaging techniques have not had sufficient resolution to distinguish between them. Here we apply the S-wave ‘receiver function’ technique to data of three permanent seismic broadband stations on the Hawaiian islands, to map the thickness of the underlying lithosphere. We find that under Big Island the lithosphere is 100–110 km thick, as expected for an oceanic plate 90–100 million years old that is not modified by a plume. But the lithosphere thins gradually along the island chain to about 50–60 km below Kauai. The width of the thinning is about 300 km. In this zone, well within the larger-scale topographic swell, we infer that the rejuvenation model⁵ (where the plume thins the lithosphere) is operative; however, the larger-scale topographic swell is probably supported dynamically.

Thermomechanical processes caused by plume–lithosphere interactions have been studied for several decades. Various models have been proposed to explain the generating mechanism of the Hawaiian swell: (1) a model in which the dynamic pressure of the plume material flowing in the asthenosphere away from the plume centre supports the uplift of the sea floor^{4,6}; (2) the thermal resetting (rejuvenation) model, in which the oceanic lithosphere is heated when it passes the hotspot, and thinning of the lithosphere causes isostatic uplift of the Hawaiian swell⁵; (3) a model in which

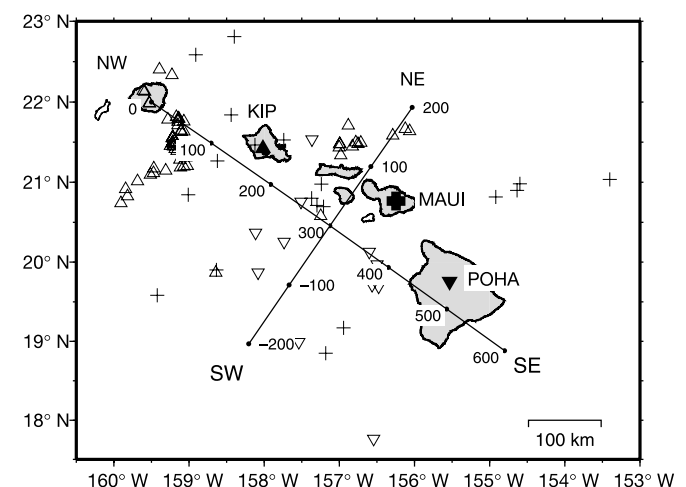


Figure 1 Location map of permanent broadband stations KIP (IRIS/GEOSCOPE), MAUI (GEOFON) and POHA (IRIS/USGS). Piercing points of S-receiver functions at 100 km depth are shown (same symbols as stations). Two lines labelled NW–SE and SW–NE indicate the two sections shown in Fig. 2 onto which the S-receiver functions are projected. Section NW–SE is along the island chain while section SW–NE is perpendicular to the chain.

extracted basaltic melt of the plume material underplates the base of the lithosphere. This residue of the plume material is compositionally lighter and has high seismic velocity^{7–9}. Different models predict different seismic thicknesses of the lithosphere. In the first model, the lithospheric thickness along the swell is relatively stable. The second model (thermal rejuvenation) predicts a thinner lithosphere when it passes the hotspot. The lithospheric thickness in the third model will be larger than normal.

Until now, only the following observations of the lithospheric thickness by means of different seismic data have been available. In studies of surface-wave dispersion between Big Island and Oahu, a thickness of 88 km was suggested for a lithosphere 80–90 million years (Myr) old¹⁰. Between Oahu and Midway, 2,000 km northwest of Oahu, the average lithospheric thickness is estimated to be 100 km for a 50–110-Myr-old lithosphere^{11,12}. In studies using converted waves, the thickness beneath Oahu was suggested to be 65–80 km (refs 13, 14), and 80 km southwest of Oahu¹⁵. From surface-wave dispersion data of an OBS (ocean bottom seismometer) array southwest of Oahu and Big Island, the average seismic structure of a 100-Myr-old oceanic lithosphere was derived and a significant velocity reduction at asthenospheric depths closer than 300 km to the island chain was found¹⁶. These facts are not sufficient to distinguish between competing models of lithosphere–plume interaction.

We have applied seismic techniques (S-receiver functions^{13,17,18}) developed for P-receiver function studies to S-to-P converted waves (see Supplementary Information for a comparison of both techniques). S-receiver functions have been calculated from data of the permanent broadband stations POHA (IRIS/USGS), MAUI (GEOFON/ GeoForschungsZentrum, Potsdam) and KIP (IRIS/ GEOSCOPE), which are located on the islands of Hawaii, Maui and Oahu, respectively. The locations of the stations POHA, MAUI and KIP and the piercing points of the S-to-P conversions at 100 km

depth are shown in Fig. 1. Figure 2 shows the stacked S-receiver functions for a section NW–SE along the island chain (Fig. 2a) and a section SW–NE perpendicular to it (Fig. 2b). The sections are formed by stacking individual traces with piercing points at depths of 100 km within a window of length 30 km. Then the window is moved by 20 km for a new summation trace. Before summation, all traces were corrected for epicentral distance or slowness. A reference slowness of 6.4 s deg^{-1} has been chosen, similar to that in many P-receiver function studies, making P-to-S and S-to-P differential times directly comparable.

To suppress oceanic noise, a 7–20 s bandpass filter is applied. We reversed the polarity of amplitudes to compare the data directly with P-receiver functions. A clear negative phase (LAB, lithosphere–asthenosphere boundary) is identified in Fig. 2a, b. Negative (shaded) amplitudes are caused by velocity reduction downwards. To confirm that this phase is caused by the LAB, we compared the S-receiver functions with conventional P-receiver functions at the same stations and found good agreements (see Supplementary Information). The data in Fig. 2 are in the time domain. These delay times (in seconds) can roughly be converted into depth (in kilometres) by multiplying all times with a factor of 9, from the IASP91 model.

Figure 2a shows clearly that along the island chain the LAB thins from Big Island (12 s or 108 km depth at profile kilometre 500) to Kauai (6 s or 54 km depth at profile kilometre 0). Perpendicular to the island chain on the SW–NE section (Fig. 2b), the LAB is at 12 s delay time (108 km deep) at profile kilometres –150 to –100 and again at profile kilometre 200, northeast of the chain. It thins to 7–8 s (63–72 km deep) at profile kilometres 50–100 which is directly beneath the island chain along the profile (Fig. 1). The NW–SE section (Fig. 2a) contains all available S-to-P data. The perpendicular SW–NE section (Fig. 2b) contains only records with piercing points southeast of Oahu, which have similar minimum LAB depths. The two profiles in Fig. 2 indicate thinning of the lithosphere from Big Island to Oahu and Kauai as imaged in Fig. 3. The end of the thinning in the northwest direction is not visible. A station on Kauai or OBS stations northwest of Kauai would be necessary to follow the continuation of the lithospheric thinning.

We obtain from our observations the following picture of the plume–lithosphere interaction. Whereas the Pacific plate overrides the Hawaiian plume with a velocity of $7\text{--}8 \text{ cm yr}^{-1}$, plume material hits the lithosphere and is dragged off. It spreads mainly in the direction of the plate motion and causes a long tail along the Hawaiian island chain. The bottom of the lithosphere is heated by

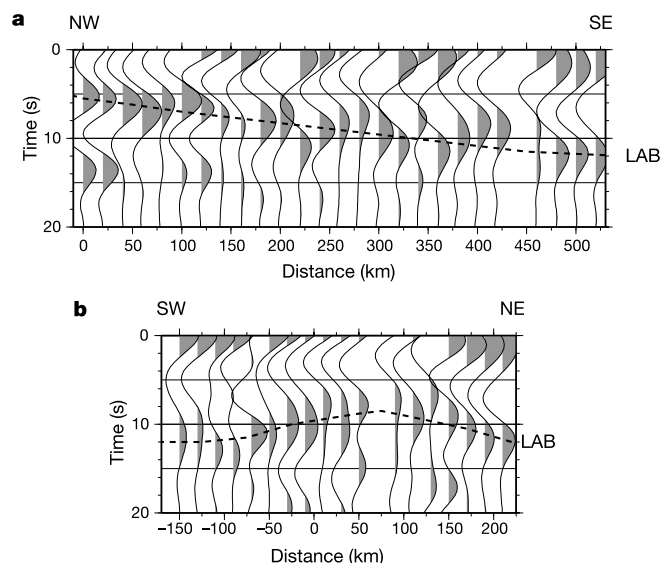


Figure 2 Stacked S-receiver functions of the three stations projected on the two profiles of Fig. 1 according to their piercing points at 100 km depth. **a**, Along the island chain (NW–SE); **b**, perpendicular to it (SW–NE). Zero time corresponds to the S-wave arrival. The timescale is reversed. The stack bin is 30 km with 10 km overlap. Slowness (or distance) move-out corrections are applied (reference slowness of 6.4 s deg^{-1}) before stacking. In section NW–SE (**a**) all S-to-P data are used that cover an up to 200-km broad sampling region at 100 km depth on each side of the section. In section SW–NE (**b**) only S–P data southeast of KIP (Oahu) are used. A 7–20 s bandpass filter is applied to suppress the noise. LAB indicates the lithosphere–asthenosphere boundary, which is a precursor of the S wave.

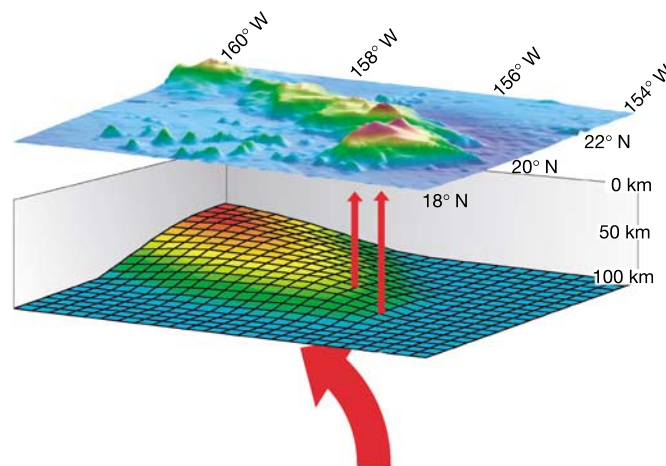


Figure 3 Sketch of the obtained lithosphere–asthenosphere boundary below the Hawaiian island chain.

the plume material from below and the increased temperature will reduce the velocity of elastic waves. In the vicinity of the present plume centre, only the lowest part of the lithosphere is heated and the lithospheric thickness does not change much with respect to that of a 90–100-Myr-old oceanic plate. As time increases, more and more material from the lower part of the lithosphere is warmed up and the seismic LAB becomes shallower. After 3–4 Myr of heating time the thinnest lithosphere of 50–60 km is achieved beneath Oahu and Kauai—a location about 400 km away from the current plume centre¹⁰. Further away, at Midway Island, which passed the plume centre about 28 Myr ago, the plume material begins to lose its temperature and the lithosphere thickens again^{11,12}.

We conclude that the Hawaiian plume has gradually reheated up to 50 km of the lower lithosphere within the last 3–4 Myr in an area of 500 × 300 km along the island chain. This means that a substantial amount of heat is trapped below the lithosphere. The lateral extent of the asthenospheric updoming is significantly smaller than the extent of the Hawaiian topographic swell. A consequence of our result could be that the entire topographic swell is caused by, for example, the dynamic support of the plume, and so rejuvenation would be contributing only in the central part. Data from longer-term broadband stations on Kauai, and on the ocean bottom northwest of Kauai, could permit imaging of the ‘re-ageing’ of the rejuvenated Hawaiian lithosphere. □

Received 27 September 2003; accepted 14 January 2004; doi:10.1038/nature02349.

1. Wilson, J. T. A possible origin of the Hawaiian island. *Can. J. Phys.* **41**, 863–868 (1963).
2. Morgan, W. J. Convective plumes in the lower mantle. *Nature* **230**, 42–43 (1971).
3. Nataf, H.-C. Seismic imaging of mantle plumes. *Annu. Rev. Earth Planet. Sci.* **28**, 391–417 (2000).
4. Ribe, N. M. & Christensen, U. R. The dynamical origin of the Hawaiian volcanism. *Earth Planet. Sci. Lett.* **171**, 517–531 (1999).
5. Detrick, R. S. & Crough, S. T. Island subsidence, hot spots, and lithospheric thinning. *J. Geophys. Res.* **83**, 1236–1244 (1978).
6. Sleep, N. H. Hotspots and mantle plumes: Some phenomenology. *J. Geophys. Res.* **95**, 6715–6736 (1990).
7. Jordan, T. H. in *The Mantle Sample: Inclusions in Kimberlites and Other Volcanics* (eds Boyd, F. R. & Meyer, H. O. A.) 1–14 (Proc. 2nd Int. Kimberlite Conference, American Geophysical Union, 1979).
8. Robinson, E. M. The topographic and gravitational expression of density anomalies due to melt extraction in the uppermost oceanic mantle. *Earth Planet. Sci. Lett.* **90**, 221–228 (1988).
9. Phipps Morgan, J., Morgan, W. J., Zhang, Y.-S. & Smith, W. H. F. Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection. *J. Geophys. Res.* **100**, 12753–12767 (1995).
10. Priestley, K. & Tilmann, F. Shear-wave structure of the lithosphere above the Hawaiian hot spot from two-station rayleigh wave phase velocity measurements. *Geophys. Res. Lett.* **26**, 1493–1496 (1999).
11. Woods, M., Leveque, J. J. & Okal, E. A. Two-station measurements of Rayleigh wave group velocity along the Hawaiian swell. *Geophys. Res. Lett.* **18**, 105–108 (1991).
12. Woods, M. T. & Okal, E. A. Rayleigh-wave dispersion along the Hawaiian Swell: a test of lithospheric thinning by the thermal rejuvenation at a hotspot. *Geophys. J. Int.* **125**, 325–339 (1996).
13. Bock, G. Long-period S to P converted waves and the onset of partial melting beneath Oahu, Hawaii. *Geophys. Res. Lett.* **18**, 869–872 (1991).
14. Li, X. *et al.* Mapping the Hawaiian plume with converted seismic waves. *Nature* **405**, 938–941 (2000).
15. Collins, J. A., Vernon, F. L., Orcutt, J. A. & Stephen, R. A. Upper mantle structure beneath the Hawaiian swell: constraints from the ocean seismic network pilot experiment. *Geophys. Res. Lett.* **29**, doi: 10.1029/2001GL013302 (2002).
16. Laske, G., Phipps Morgan, J. & Orcutt, J. A. First results from the Hawaiian SWELL pilot experiment. *Geophys. Res. Lett.* **26**, 3397–3400 (1999).
17. Faber, S. & Müller, G. Sp phases from the transition zone between the upper and lower mantle. *Bull. Seismol. Soc. Am.* **70**, 487–508 (1980).
18. Farra, V. & Vinnik, L. Upper mantle stratification by P and S receiver functions. *Geophys. J. Int.* **141**, 699–712 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Asch and J. Mechie for their aid. We also thank the people of the Haleakala National Park for supporting our station MAUI. This work has been supported by the Deutsche Forschungsgemeinschaft within the ICDP project and by the GeoForschungsZentrum, Potsdam. Waveform data have been provided by the IRIS, GEOSCOPE and GEOFON data centres.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (kind@gfz-potsdam.de).

Iron corrosion by novel anaerobic microorganisms

Hang T. Dinh¹, Jan Kuever^{1,2}, Marc Mußmann¹, Achim W. Hassel³, Martin Stratmann³ & Friedrich Widdel¹

¹Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

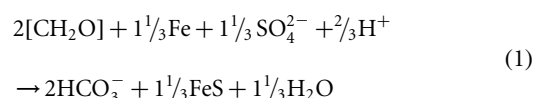
²Institute for Material Testing, Paul-Feller-Straße 1, 28199 Bremen, Germany

³Max Planck Institute for Iron Research, Max-Planck-Straße 1, 40237 Düsseldorf, Germany

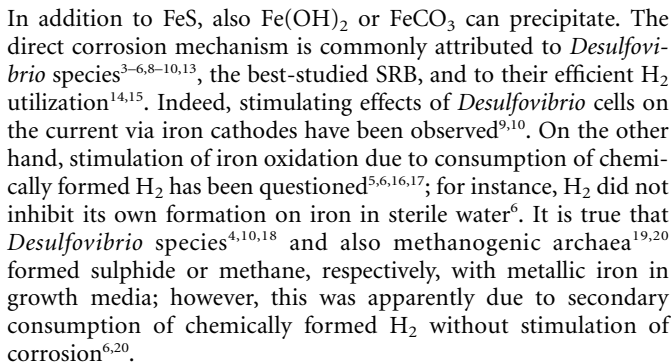
Corrosion of iron presents a serious economic problem. Whereas aerobic corrosion is a chemical process¹, anaerobic corrosion is frequently linked to the activity of sulphate-reducing bacteria (SRB)^{2–6}. SRB are supposed to act upon iron primarily by produced hydrogen sulphide as a corrosive agent^{3,5,7} and by consumption of ‘cathodic hydrogen’ formed on iron in contact with water^{2–6,8}. Among SRB, *Desulfovibrio* species—with their capacity to consume hydrogen effectively—are conventionally regarded as the main culprits of anaerobic corrosion^{2–6,8–10}; however, the underlying mechanisms are complex and insufficiently understood. Here we describe novel marine, corrosive types of SRB obtained via an isolation approach with metallic iron as the only electron donor. In particular, a *Desulfobacterium*-like isolate reduced sulphate with metallic iron much faster than conventional hydrogen-scavenging *Desulfovibrio* species, suggesting that the novel surface-attached cell type obtained electrons from metallic iron in a more direct manner than via free hydrogen. Similarly, a newly isolated *Methanobacterium*-like archaeon produced methane with iron faster than do known hydrogen-using methanogens, again suggesting a more direct access to electrons from iron than via hydrogen consumption.

Some 10% of all corrosion damages to metals and non-metals may result from microbial activities¹¹. A significant process in this respect is the anaerobic corrosion of iron or steel, for instance in oil technology^{3,6,11}. The primary dissolution ($\text{Fe} \rightleftharpoons \text{Fe}^{2+} + 2\text{e}^-$; $E^0 = -0.44 \text{ V}$) can in principle be driven by numerous oxidants. In oxic humid surroundings, oxidation by O_2 ($E_{\text{pH}7}^0 = +0.82 \text{ V}$) yields rust¹. In anoxic surroundings, H^+ ions from water yield H_2 ($E_{\text{pH}7}^0 = -0.41 \text{ V}$). Of the sequential steps ($2\text{e}^- + 2\text{H}^+ \rightarrow 2\text{H}_{(\text{adsorbed})} \rightarrow \text{H}_{2(\text{adsorbed})} \rightarrow \text{H}_{2(\text{aqueous})}$), the combination of the H atoms is presumably rate-limiting¹² and is a main reason for the slowness of iron oxidation in anoxic sterile water ($\text{Fe} + 2\text{H}_2\text{O} \rightarrow \text{Fe}^{2+} + \text{H}_2 + 2\text{HO}^-$). However, sulphate-reducing bacteria (SRB) promote the anaerobic oxidation of iron^{2–6}. Their activity often occurs in biofilms and tends to pit the iron^{2,3,13}. An indirect and a direct corrosion mechanism may be distinguished^{3,5}: these may occur simultaneously at different extents, depending on the load of waters with biodegradable organic compounds.

The indirect mechanism is a chemical attack by hydrogen sulphide ($\text{Fe} + \text{H}_2\text{S} \rightarrow \text{FeS} + \text{H}_2$) which is faster than that by water^{3,5,7} and also promotes so-called hydrogen embrittlement of the metal^{11,12} (see also <http://www.corrosion-doctors.org>). Because SRB commonly use organic compounds (shown here as $[\text{CH}_2\text{O}]$) and often also H_2 for sulphate reduction, the net reaction of indirect corrosion (for complete carbon oxidation by SRB communities) can be written as:



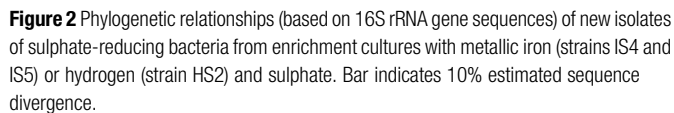
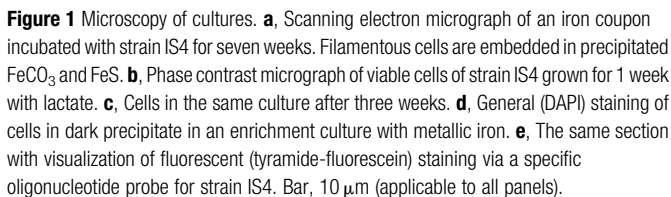
In the direct mechanism according to the depolarization theory (see Supplementary Information)⁸, SRB are supposed to stimulate corrosion by scavenging ‘cathodic hydrogen’ or a ‘hydrogen film’

$$4\text{Fe} + \text{SO}_4^{2-} + 4\text{H}_2\text{O} \rightarrow \text{FeS} + 3\text{Fe}^{2+} + 8\text{HO}^- \quad (2)$$


Two representative strains, IS4 and IS5, were isolated from iron-

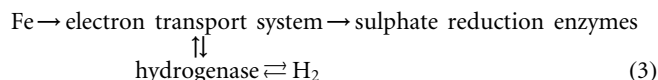
Probing of the enrichment culture with a fluorescent 16S ribosomal RNA-targeted oligonucleotide specific for strain IS4 revealed high numbers of precipitate-associated cells with shape similar to that of strain IS4 (Fig. 1d, e); they represented the majority of the detectable cells. A common oligonucleotide probe for *Desulfovibrio* species²² did not hybridize.

For direct measurement of corrosiveness, we followed sulphate reduction or methanogenesis with metallic iron as the only electron donor by the newly isolated as well as by authenticated SRB and methanogenic archaea, respectively (Fig. 3a, b). The authenticated species are known hydrogen utilizers. Sulphate reduction to sulphide by strain IS4 with iron was fast, much as in the enrichment culture. Sulphate reduction gradually slowed down, but became faster again if fresh iron was added (not shown), suggesting that formed crusts (Fig. 1a) act as process barriers. Sulphate reduction by strain IS5 was slower than by strain IS4, but faster than by strain HS2 and the authenticated *D. salexigens* and *D. vulgaris*. Sulphate reduction rates of the latter three were in agreement with a secondary consumption of chemically formed H_2 (1 mol sulphate requiring 4 mol H_2). Methanogenesis with iron was also more pronounced with the new strain IM1 than with the authenticated *Methanococcus maripaludis* (Fig. 3b), *Methanogenium organophilum* and *Methanosarcina mazei* (not shown). The rate of methanogenesis with iron by the authenticated species was again in accordance with a secondary consumption of chemically formed H_2 (1 mol CH_4 requiring 4 mol H_2).



Because the speed of sulphate reduction by strains IS4 and IS5 and of methanogenesis by strain IM1 with iron cannot be explained by mere consumption of the chemically formed hydrogen, these organisms must obtain reducing equivalents more efficiently. In cultures of strain IS4 with iron, H_2 accumulated to high levels (Fig. 3c) and hence cannot be the rate-limiting intermediate; this was not observed in incubations with strains IS5, HS2 and the authenticated *Desulfovibrio* species. An efficient use of metallic iron for sulphate reduction (or methanogenesis) would be possible by an electron uptake via a cell-surface-associated redox-active component, a principle known from aerobic iron(II)-oxidizing bacteria²³. The inverse process, a delivery of electrons (from organic electron donors) by cells to external solid or dissolved matter occurs in iron(III)-reducing bacteria²⁴. Following hydrogen formation assays with cell fractions of *D. vulgaris*, a cytochrome in the outer membrane has been suggested to participate in iron corrosion; an electron flow to the sulphate reduction enzymes via two types of periplasmic hydrogenase and H_2 as a direct intermediate was proposed, according to: $Fe \rightarrow \text{cytochrome} \rightarrow \text{hydrogenase(1)} \rightarrow H_2 \rightarrow \text{hydrogenase(2)} \rightarrow \text{electron transport system} \rightarrow \text{sulphate reduction enzymes}$ ¹⁷ (arrows indicate electron flow). In our study, however, the same *D. vulgaris* did not reveal a corrosive potential like strains IS4 and IS5. Nevertheless, the assumption of a cell-surface-associated redox component is an obvious hypothesis to explain how the new isolates can circumvent the slow chemical formation of H_2 (ref. 12) to obtain their reducing equivalents. The

pronounced H_2 formation during growth of strain IS4 with metallic iron does not necessarily indicate that H_2 is a direct biochemical intermediate connecting two hydrogenases¹⁷. H_2 may just as well be formed via a branch according to:



as a result of an imbalance between electron donation by fresh iron and electron consumption by sulphate reduction (see also Supplementary Fig. 3). The hydrogenase may otherwise function in growth with external hydrogen. Detailed models of electron flow are at present not possible because knowledge of the topology and function of redox proteins in various SRB¹⁴ (and methanogens²⁵) is insufficient.

The abundance of cells resembling strain IS4 in our enrichment culture with iron and the effective iron-dependent sulphate reduction by this strain suggests a thus-far overlooked involvement of such or physiologically similar SRB in anaerobic corrosion. Proof of this assumption requires the examination of technical plants with corroding iron or steel. The natural significance of the capacity for using metallic iron as electron donor is unknown. Apart from rare meteorites, metallic iron is a technical product and is thus a very 'recent' growth substrate. One may speculate that iron-using anaerobes can also obtain electrons in cell contact with certain microorganisms. □

Methods

Enrichment, isolation and cultivation

Marine sediment was collected near Wilhelmshaven, North Sea. Cultures were grown at 28 °C in anoxic seawater (marine organisms) or freshwater (*D. vulgaris*) medium with 28 mM sulphate (0.1 mM for methanogens) under $N_2 + CO_2$ (90/10, vol./vol.)¹⁵. If indicated, 1 mM sodium acetate was added as organic carbon source. Iron granules (99.8% Fe, size 2 mm, 20 g per 100 ml) or mild steel coupons (1 mm thickness, fitted to tubes or bottles) were added as the electron source. For subculturing, 10% of the culture liquid and part of the iron specimens were transferred every seven weeks. Parallel enrichment cultures were carried out with $H_2 + CO_2$ (+1 mM acetate) without metallic iron.

For isolation of enriched organisms, precipitates from the iron surface were homogenized anaerobically and sequentially diluted in medium with iron granules. The highest dilutions showing sulphate reduction or methanogenesis were again diluted. SRB were finally diluted in anoxic agar¹⁵ with a mixture of lactate, propionate, butyrate, pyruvate, ethanol (each 2 mM) and hydrogen. Colonies were transferred to liquid medium.

Authenticated strains were from the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany).

Chemical analyses

Sulphate was determined micro-gravimetrically as $BaSO_4$ precipitated from 1-ml samples with the same volume of 0.2 M $BaCl_2$ and 0.2 M HCl. CH_4 and H_2 were quantified on a gas chromatograph with a flame ionization or thermal conductivity detector, respectively. Minerals on the iron surface were analysed by X-ray photoelectron spectroscopy.

Analyses of 16S rRNA genes

DNA was extracted with the DNeasy Tissue kit (Qiagen). Nearly full-length (1,500 bp) 16S rRNA gene sequences from SRB were amplified using general bacterial primers²⁶. A partial (about 1,000 bp) 16S rRNA gene sequence of the methanogenic isolate was amplified with archaeal primers²⁷. Sequences were analysed using the ARB program²⁸, and a phylogenetic tree was constructed via maximum-parsimony, neighbour-joining and maximum-likelihood analyses. The methanogenic strain IM1 was positioned in the tree according to parsimony criteria without affecting the overall topology.

Specific cell hybridization and unspecific staining

Precipitates scratched from corroded iron surfaces were fixed for 12 h at 4 °C in 4% formaldehyde, washed twice with PBS (10 mM sodium phosphate, pH 7; 130 mM NaCl), stored in PBS-ethanol (1:1) at -20 °C, and collected on polycarbonate filters (0.2 µm pores; Millipore). A horseradish peroxidase-labelled probe (5'-CTCCTCTGCTGCTAGCT-3') was specifically designed and synthesized (ThermoHybaid) for strain IS4. After hybridization at 35 °C in the presence of 55% (vol./vol.) formamide and washing, the probe was reacted with tyramide-fluorescein²⁹. DAPI (4',6-diamidino-2-phenylindole) was used for general cell staining.

Remarks on redox potentials and equations

Indicated redox potentials are versus standard hydrogen electrode ($E^0 = 0$ V). The Fe^{2+}/Fe redox potential in cultures and sea water is much more negative than under standard conditions (-0.44 V); whereas Fe remains in standard state, salts and

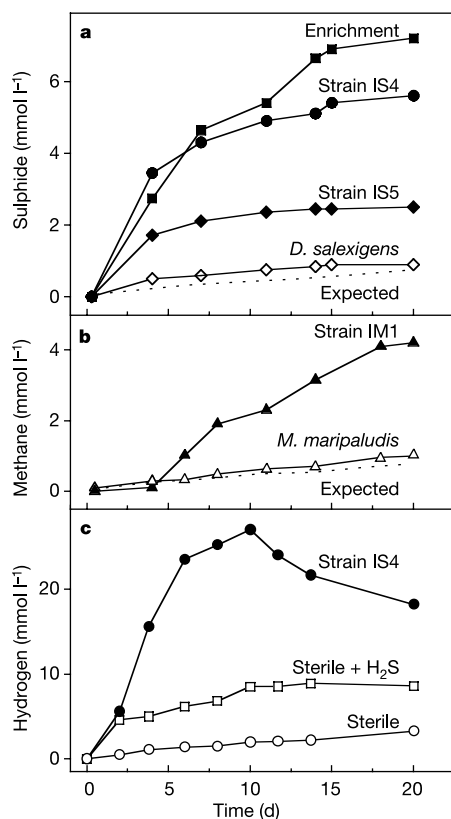


Figure 3 Incubation experiments with iron granules (30 g in 150 ml medium) as sole electron donor. **a**, Sulphide formation (via sulphate consumption) by strains IS4, IS5 and *Desulfovibrio salexigens*, and the original enrichment culture. **b** Methane formation by strain IM1 and *Methanococcus maripaludis*. **c**, Hydrogen formation by strain IS4 and in sterile iron incubations without or with hydrogen sulphide (4 mM). In the latter, hydrogen production became slower after binding of free sulphide as FeS . **a**, **b**, Sulphate reduction or methanogenesis expected from mere consumption of chemically formed hydrogen is indicated by dotted lines.

precipitation as sulphide and carbonate decrease the Fe^{2+} activity⁵ such that redox potentials around -0.53 V are realistic. Proton reduction to H_2 ($E_{\text{H}^+/\text{H}_2}^0 = -0.41$ V) on metallic iron in such surroundings would thus be far from the thermodynamic equilibrium.

Equation (1) is based on the equations for sulphate reduction with the organic compound ($2[\text{CH}_2\text{O}] + \text{SO}_4^{2-} \rightarrow 2\text{HCO}_3^- + \text{H}_2\text{S}$) and hydrogen ($4\text{H}_2 + \text{SO}_4^{2-} + 2\text{H}^+ \rightarrow \text{H}_2\text{S} + 4\text{H}_2\text{O}$). If SRB reduce 1 mol SO_4^{2-} with an organic compound to 1 mol H_2S , the latter yields 1 mol H_2 upon chemical reaction with Fe to FeS. Use of H_2 for further sulphate reduction yields $1/4$ mol H_2S which leads to $1/4$ mol H_2 . Continuation ad infinitum leads to a total of $1^{1/3}$ (sum of infinite row $1 + 1/4 + 1/16$ etc.) mol H_2S that attacks the iron. (For other remarks on reactions in corrosion see Supplementary Information.)

Received 25 April; accepted 23 December 2003; doi:10.1038/nature02321.

- Uhlir, H. H. *Corrosion and Corrosion Control* 3rd edn (Wiley, New York, 1985).
- Hamilton, W. A. Microbially influenced corrosion as a model system for the study of metal microbe interactions: a unifying electron transfer hypothesis. *Biofouling* **19**, 65–76 (2003).
- Lee, W., Lewandowski, Z., Nielsen, P. H. & Hamilton, W. A. Role of sulfate-reducing bacteria in corrosion of mild steel: a review. *Biofouling* **8**, 165–194 (1995).
- Pankhania, I. P. Hydrogen metabolism in sulphate-reducing bacteria and its role in anaerobic corrosion. *Biofouling* **1**, 27–47 (1988).
- Widdel, F. in *Biotechnology Focus 3* (eds Finn, R. K. et al.) 277–318 (Hanser, Munich, 1992).
- Cord-Ruwisch, R. in *Environmental Microbe-Metal Interaction* (ed. Lovley, D. R.) 159–173 (ASM Press, Washington, DC, 2000).
- Costello, J. A. Cathodic depolarization by sulphate-reducing bacteria. *S. Afr. J. Sci.* **70**, 202–204 (1974).
- von Wolzogen Kuehr, C. A. H. & van der Vlugt, I. S. The graphitization of cast iron as an electrochemical process in anaerobic soil. *Water* **18**, 147–165 (1934).
- Booth, G. H. & Tiller, A. K. Cathodic characteristic of mild steel in suspension of sulphate-reducing bacteria. *Corros. Sci.* **8**, 583–600 (1968).
- Pankhania, I. P., Moosavi, A. N. & Hamilton, W. A. Utilization of cathodic hydrogen by *Desulfovibrio vulgaris* (Hildenborough). *J. Gen. Microbiol.* **132**, 3357–3365 (1986).
- Iverson, W. P. & Olson, G. J. in *Current Perspectives in Microbial Ecology* (eds Klug, M. J. & Reddy, C. A.) 623–627 (ASM, Washington, DC, 1984).
- Bockris, J. O'M. & Reddy, A. K. N. *Modern Electrochemistry* Vol. 2 (Plenum, New York, 1970).
- Beech, I. B. et al. Study of parameters implicated in the biodeterioration of mild steel in the presence of different species of sulphate-reducing bacteria. *Int. Biodeter. Biodegrad.* **34**, 289–303 (1994).
- Rabus, R., Hansen, T. & Widdel, F. in *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community* (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) (Springer, New York, 2000).
- Widdel, F. & Bak, F. in *The Prokaryotes* 2nd edn Vol. 6 (eds Balows, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H.) 3352–3378 (Springer, New York, 1992).
- Hardy, J. A. Utilisation of cathodic hydrogen by sulphate-reducing bacteria. *Br. Corros. J.* **18**, 190–193 (1983).
- Laishley, E. J. & Bryant, R. D. in *Biochemistry and Physiology of Anaerobic Bacteria* (eds Ljungdahl, L. G., Adams, M. W., Barton, L. L., Ferry, J. G. & Johnson, M. K.) 252–260 (Springer, New York, 2003).
- Cord-Ruwisch, R. & Widdel, F. Corroding iron as a hydrogen source for sulphate reduction in growing cultures of sulphate-reducing bacteria. *Appl. Microbiol. Biotechnol.* **25**, 169–174 (1986).
- Daniels, L., Belay, N., Rajagopal, B. S. & Weimer, P. J. Bacterial methanogenesis and growth from CO_2 with elemental iron as the sole source of electrons. *Science* **237**, 509–511 (1987).
- Deckena, S. & Blotvogel, K.-H. Fe^0 -oxidation in the presence of methanogenic and sulphate-reducing bacteria and its possible role in anaerobic corrosion. *Biofouling* **5**, 287–293 (1992).
- Schlegel, H. G. *General Microbiology* 7th edn (Cambridge Univ. Press, Cambridge, 1993).
- Manz, W., Eisenbrecher, M., Neu, T. R. & Szwedzyk, U. Abundance and spatial organization of Gram-negative sulfate-reducing bacteria in activated sludge investigated by in situ probing with specific 16S rRNA targeted oligonucleotides. *FEMS Microbiol. Ecol.* **25**, 43–61 (1998).
- Appia-Ayme, C., Guiliani, N., Ratouchniak, J. & Bonnefoy, V. Characterization of an operon encoding two c-type cytochromes, an aa₃-type cytochrome oxidase, and rusticyanin in *Thiobacillus ferrooxidans* ATCC 33020. *Appl. Environ. Microbiol.* **65**, 4781–4787 (1999).
- Bond, D. R. & Lovley, D. R. Electricity production by *Geobacter sulfurreducens* attached to electrodes. *Appl. Environ. Microbiol.* **69**, 1548–1555 (2003).
- Deppenmeier, U. The unique biochemistry of methanogenesis. *Prog. Nucleic Acid Res. Mol. Biol.* **71**, 223–283 (2002).
- Muyzer, G., Teske, A., Wirsén, C. O. & Jannasch, H. W. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch. Microbiol.* **164**, 165–172 (1995).
- Huber, H. et al. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).
- Ludwig, W. et al. ARB: a software environment for sequence data. (Department of Microbiology, Technical Univ. Munich, 2002); available at (<http://www.arb-home.de/>).
- Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68**, 3094–3101 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Nauhaus, M. Nellesen and G. Herz for technical help. This study was supported by the German Academic Exchange Service, the Fonds der Chemischen Industrie, and the Max Planck Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.W. (fwiddel@mpi-bremen.de). The nucleotide sequences have been deposited at EMBL GenBank under accession numbers AY274444 (strain HS2), AY274449 (strain IS5), AY274450 (strain IS4) and AY274451 (strain IM1).

Conventional taxonomy obscures deep divergence between Pacific and Atlantic corals

Hironobu Fukami^{1,2}, Ann F. Budd³, Gustav Paulay⁴, Antonio Solé-Cava⁵, Chaolun Allen Chen⁶, Kenji Iwao⁷ & Nancy Knowlton^{1,2}

¹Smithsonian Tropical Research Institute, Naos Marine Laboratory, Box 2072, Balboa, Republic of Panama

²Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0202, USA

³Department of Geoscience, University of Iowa, Iowa City, Iowa 52242, USA

⁴Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611-7800, USA

⁵Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

⁶Institute of Zoology, Academia Sinica, Nankang, Taipei 115, Taiwan

⁷Akajima Marine Science Laboratory, Zamami-son, Okinawa 901-3311, Japan

Only 17% of 111 reef-building coral genera and none of the 18 coral families with reef-builders are considered endemic to the Atlantic, whereas the corresponding percentages for the Indo-west Pacific are 76% and 39%^{1,2}. These figures depend on the assumption that genera and families spanning the two provinces belong to the same lineages (that is, they are monophyletic). Here we show that this assumption is incorrect on the basis of analyses of mitochondrial and nuclear genes. Pervasive morphological convergence at the family level has obscured the evolutionary distinctiveness of Atlantic corals. Some Atlantic genera conventionally assigned to different families are more closely related to each other than they are to their respective Pacific 'congeners'. Nine of the 27 genera of reef-building Atlantic corals belong to this previously unrecognized lineage, which probably diverged over 34 million years ago. Although Pacific reefs have larger numbers of more narrowly distributed species, and therefore rank higher in biodiversity hotspot analyses³, the deep evolutionary distinctiveness of many Atlantic corals should also be considered when setting conservation priorities.

We tested the assumption of inter-oceanic monophyly by examining the relationships of Atlantic and Pacific members of the Faviidae and Mussidae, ecologically dominant families that comprise one-third of all reef-building or zooxanthella-containing coral genera^{1,2}. We analysed mitochondrial cytochrome B (*cytB*) and cytochrome oxidase I (*COI*) genes as well as parts of two exons from a nuclear β -tubulin gene. We sequenced representatives of more than half the genera in both families, including most of the Caribbean species, as well as taxa in related families.

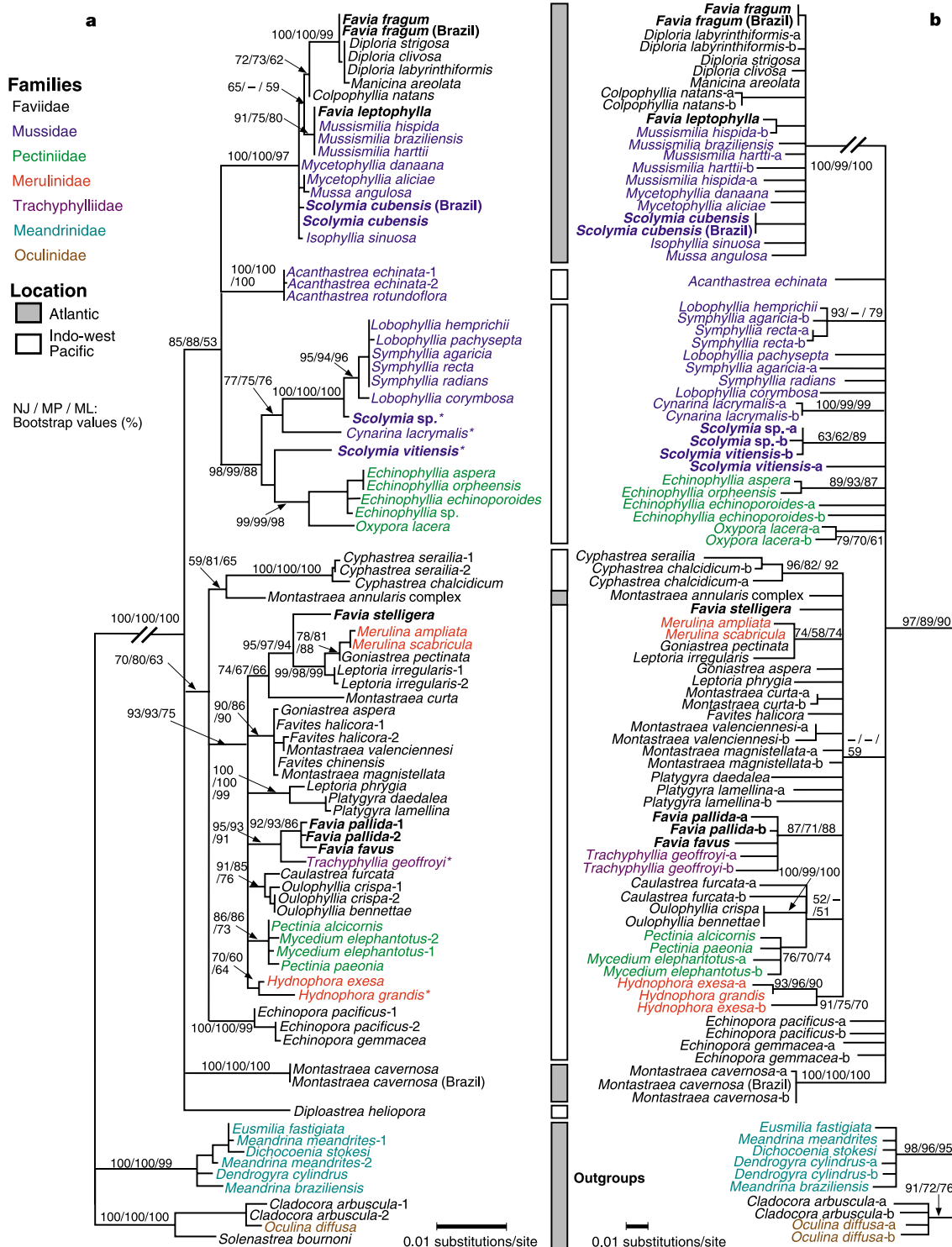
Unexpectedly, most Atlantic lineages conventionally assigned to the Faviidae and Mussidae are not distributed within the more numerous Pacific lineages of these 'families', but instead represent a well-defined clade (Fig. 1). Notably, the Atlantic 'faviid' *Favia* and the Atlantic 'mussid' *Scolymia* are more closely related to each other than they are to their respective 'congeners' in the Pacific (taxa in bold in Fig. 1). The only Atlantic corals closely related to the group containing Pacific 'faviids' belong to the clearly polyphyletic 'genus' *Montastraea*; seven other genera in this group that are now restricted to the Pacific once had Atlantic representatives (on the basis of conventional taxonomy)⁴.

Moreover, the distinctiveness of this previously unrecognized Atlantic clade is greater than that of several conventionally recognized families that are restricted to the Indo-west Pacific. Specifically, the Merulinidae and Pectiniidae (which in our analyses are polyphyletic), as well as the monotypic Trachyphyllidae, are nested within and are more closely related to Pacific 'faviids' and 'mussids'

than the Atlantic clade is related to any of these taxa. In other words, the Atlantic clade clearly deserves recognition at the family level (Mussidae Ortmann 1890 is the senior available family name), whereas several currently recognized Pacific families probably do not.

These two conclusions are supported by mitochondrial and

nuclear genes analysed by a variety of methods (Fig. 1, and Supplementary Information). Previous molecular analyses of corals^{5–8} did not come to comparable conclusions, either because they did not simultaneously analyse large numbers of Pacific and Atlantic corals in these groups or did not have adequate resolution to detect these patterns.



sequences of different clones from the same colony. Branches with less than 50% bootstrap values were collapsed; values for major nodes are indicated. Four species in **a** are missing in **b** (see Methods).

NATURE | VOL 427 | 26 FEBRUARY 2004 | www.nature.com/nature

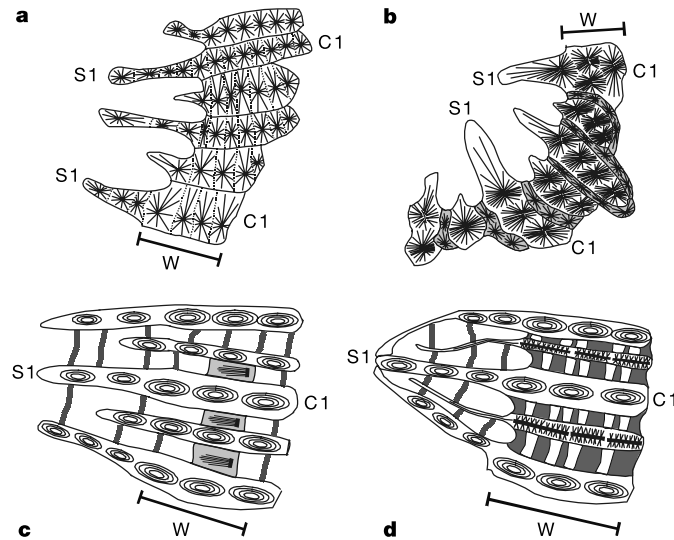


Figure 2 Illustrations of wall structures based on thin sections. **a**, *Favia fragum*, walls formed by septal thickening (septothechal). **b**, *Favia stelligera*, walls formed by insertion of skeletal structures with trabecular centres (trabeculothechal). **c**, *Mycetophyllia aliciae*, walls composed mostly of dissepiments (parathechal) with some trabeculothechal elements. **d**, *Symphyllia recta*, walls composed of dissepiments thickened by sclerenchyme.

Abbreviations: W, wall; S1, primary septum; C1, primary costa. Trabecular centres are points completely surrounded by radiating spokes (concentric ellipses represent trabecular fans). Trabeculothechal elements (**b**, **c**) are lightly shaded; dissepiments (**c**, **d**) are dark.

Our results are unexpected given the long-term taxonomic stability of zooxanthellate/reef-building scleractinian coral families. Of the 18 families currently recognized², all except three had been described by 1902, and the biogeographic coherence of the Mussidae and Faviidae has never previously been questioned^{1,2,9}. This suggests that new morphological characters and approaches must be identified for both living and fossil taxa^{8,10}. A full morphological analysis is beyond the scope of this paper, but several characters largely ignored by previous workers substantiate several aspects of our molecular phylogeny. For example, the corallite walls of Atlantic *Favia fragum* and other examined Atlantic 'faviids' except *Montastraea* are formed by septal thickening (Fig. 2a), whereas the walls of the Pacific *F. stelligera* (Fig. 2b) and other examined relatives are constructed by the insertion of skeletal structures with distinct trabecular centres between the costae. Atlantic 'mussids' have corallite walls formed partly by thin dissepiments and by additional fibrous structures inserted between the costae that lack trabecular centres (Fig. 2c), whereas Pacific 'mussids' have walls composed of dissepiments and thickened by sclerenchyme (Fig. 2d). Pacific members of the 'genus' *Scolymia* are more like other Pacific 'mussids' in having corallite centres with primarily lamellar linkage, whereas Atlantic *Scolymia* resemble Atlantic *Mussa* in having centres with trabecular linkage, a fact noted decades ago¹¹ but deemed insignificant in later taxonomic treatises.

This Atlantic clade represents one-third of the 27 genera of zooxanthellate or reef-building corals of the region. Its timing is difficult to specify with precision because our results imply that previous generic assignments of fossil taxa need to be re-examined. On the basis of the fossil record of the most morphologically distinctive genera (for example *Colpophyllia*, which dates back to the Eocene epoch (34–55 Myr ago)), the dominant Atlantic and Pacific lineages probably diverged long before the closure of the Tethyan connection between the tropical Indo-west Pacific and Atlantic in the early Miocene epoch (about 16–24 Myr ago), the time traditionally associated with the breakdown of connections between coral communities from these two provinces^{4,12,13}.

These new phylogenies help to make sense of previously noted discrepancies in patterns of molecular evolution for corals. It had

been noted¹⁴ that divergences between *F. fragum* and *Diploria* and among Pacific members of the 'families' Merulinidae, Pectiniidae and Faviidae are lower than that observed between the 'congeners' *M. annularis* and *M. cavernosa*, observations that are completely compatible with molecular phylogenies presented here. Our results also strengthen the case for unusually slow rates of molecular evolution in corals, because species widely separated in our phylogeny exhibit very low levels of genetic divergence (for example, *M. annularis* and *M. cavernosa*, with a *COI* divergence of only 2.4%¹⁴).

Preservation of evolutionarily distinct lineages is an important aspect of conservation¹⁵, but phylogenies have had little role in marine conservation except in the context of 'living fossils' such as the coelacanth. The evolutionary distinctiveness of these Atlantic corals indicates that factors in addition to diversity and endemism at the species level³ need to be considered in planning conservation actions, especially in view of the precarious state of most Caribbean reefs^{16,17}. □

Methods

Collections and DNA extraction

We collected 15 cm × 15 cm samples from the following locations: Bocas del Toro, Republic of Panama (Caribbean coast); the Abrolhos Archipelago, Brazil; Aka Island, Okinawa, Japan; Penghu Island, Taiwan; and islands of the Republic of Palau (see Supplementary Information for locations of skeletal vouchers and availability of photographs). Total DNA was extracted from tissue stored in modified guanidine solution (4 M guanidine thiocyanate, 0.1% *N*-lauroyl sarcosine sodium, 10 mM Tris-HCl pH 8, 0.1 M 2-mercaptoethanol) by conventional extraction with phenol/chloroform (1:1) and precipitation with ethanol. Alternatively, tissues were stored in 95% ethanol, and total DNA was extracted with a DNA isolation kit (Gentra Systems) after tissues had been homogenized in liquid nitrogen.

DNA sequencing

Coral-specific primers for mitochondrial *cytB* and *COI* were designed on the basis of the complete mitochondrial genome sequences of members of the *Montastraea annularis* complex (H.F. and N.K., manuscript in preparation). The entire *cytB* gene was amplified by polymerase chain reaction (PCR) with primers MCytbF (5'-GTT GCT AGT AAT TTG GAT TG-3') and MCytbR (5'-CAA ACC CAC CAA GCT TAA TA-3'). Half the *COI* gene was amplified with primers MCOIF (5'-TCT ACA AAT CAT AAA GAC ATA GG-3') and MCOIR (5'-GAG AAA TTA TAC CAA AAC CAG G-3'). The protocol for amplifications was 94 °C for 120 s, followed by 30 cycles at 94 °C for 45 s, 60 °C for 45 s and 72 °C for 90 s, ending with a final phase of 72 °C for 5 min. PCR products were then directly

sequenced as described in ref. 18. Internal primers McytbseqF (5'-ATT GAC TAT GGC GAC CGC TTT T-3') and McytbseqR (5'-GAA TAA AAT TCT CTG CGT CTC C-3') for *cytB* were used in addition to PCR primers.

Primers to amplify the β -tubulin gene (intron and exon regions), designed on the basis of sequence data of *Montastraea faveolata*¹⁹, were TubulinF (5'-GCA TGG GAA CGC TCC TTA TTT-3') and TubulinR (5'-ACA TCT GTT GAG TGA GTT CTG-3'). They amplify a region corresponding to amino acid positions 144–299 within exon 4 of the human and *Drosophila* β -tubulin gene; the beginning of the intron corresponds to position 247, and the flanking exons have 99% amino acid similarity to the corresponding vertebrate sequence. Depending on the genus, one, two or three bands of about 600 bases, 1.0–1.5 kilobases (kb) and more than 2.0 kb were amplified by PCR with the above-described protocol. The size difference between bands was due to the length of the intron. Because most genera had a 1.0–1.5-kb band, this was used for phylogenetic analyses. *Diploastrea* and *Solenastrea* had only the 600-base-pair (bp) band, and no bands could be amplified for *Acanthastrea rotundoflora* and *Favites chinensis*; consequently these four taxa were not analysed for β -tubulin. Amplified fragments of the β -tubulin gene were separated by agarose electrophoresis, cloned with the pGEM-T System (Promega) and sequenced for both strands. At least five clones obtained from each of two independent PCRs were analysed. If only one sequence occurred more than once, this sequence was used in the phylogenetic analyses; otherwise the two most abundant sequences were used.

DNA phylogenetic analyses

Only the exon regions of the β -tubulin gene (444 bp) were analysed, because the intron was too variable for alignment. Phylogenetic analyses were performed with PAUP^{*20}. DNA sequences of the entire *cytB* gene and the *COI* gene excluding the third codon position (total length 1,557 bases) were combined on the basis of nucleotide saturation analyses and the incongruence length difference test. Phylogenetic trees were constructed on the basis of neighbour-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML) methods with the use of PAUP^{*}. The NJ analysis was done with a two-parameter model²¹. In MP and ML analyses, heuristic searches with TBR branch swapping and 25 random additions of taxa were performed. For ML analysis, we used Modeltest²² to find an appropriate model of evolution. For mitochondrial genes the K81uf model²³ with gamma parameter (G) and proportion of invariable positions (I) was chosen. For β -tubulin we chose the TrN model²⁴ with G and I. Bootstrapping was used to evaluate support for trees (1,000 replicates for NJ and MP; 300 bootstraps with the fast-stepwise heuristic search for ML).

Received 16 December 2003; accepted 13 January 2004; doi:10.1038/nature02339.

- Veron, J. E. N. *Coral in Space and Time* (UNSW Press, Sydney, 1995).
- Veron, J. E. N. *Coral of the World* (Australian Institute of Marine Science, Townsville, 2000).
- Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284 (2002).
- Budd, A. F. Diversity and extinction in the Cenozoic history of Caribbean reefs. *Coral Reefs* **19**, 25–35 (2000).
- Romano, S. L. & Cairns, S. D. Molecular phylogenetic hypotheses for the evolution of scleractinian corals. *Bull. Mar. Sci.* **67**, 1043–1068 (2000).
- Veron, J. E. N., Odorico, D. M., Chen, C. A. & Miller, D. J. Reassessing evolutionary relationships of scleractinian corals. *Coral Reefs* **15**, 1–9 (1996).
- Chen, C. A., Wallace, C. C. & Wolstenholme, J. Analysis of the mitochondrial 12S rRNA gene supports a two-clade hypothesis of the evolutionary history of scleractinian corals. *Mol. Phylog. Evol.* **23**, 137–149 (2002).
- Cuif, J.-P., Lecointre, G., Perrin, C., Tillier, A. & Tillier, S. Patterns of septal biomineralization in Scleractinia compared with their 28S rRNA phylogeny: a dual approach for a new taxonomic framework. *Zool. Scripta* **32**, 459–473 (2003).
- Vaughan, T. W. & Wells, J. W. Revision of the suborders, families, and genera of the Scleractinia. *Geol. Soc. Am. Spec. Pap.* **44**, 1–363 (1943).
- Stolarski, J. & Roniewicz, E. Towards a new synthesis of evolutionary relationships and classification of Scleractinia. *J. Paleontol.* **75**, 1090–1108 (2001).
- Wells, J. W. The recent solitary mussels scleractinian corals. *Zool. Meded.* **39**, 375–384 (1964).
- Paulay, G. in *Life and Death of Coral Reefs* (ed. Birkeland, C.) 298–353 (Chapman & Hall, New York, 1997).
- Wallace, C. C. Journey to the heart of the centre—origins of high marine faunal diversity in the central Indo-Pacific from the perspective of an acropologist. *Proc. 9th Int. Coral Reef Symp.* **1**, 33–39 (2002).
- Shearer, T. L., van Oppen, M. J. H., Romano, S. L. & Wörheide, G. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Mol. Ecol.* **11**, 2475–2487 (2002).
- Mace, G. M., Gittleman, J. L. & Purvis, A. Preserving the tree of life. *Science* **300**, 1707–1709 (2003).
- Pandolfi, J. M. *et al.* Global trajectories of the long-term decline of coral reef ecosystems. *Science* **301**, 955–958 (2003).
- Gardner, T. A., Côté, I. M., Gill, J. A., Grant, A. & Watkinson, A. R. Long-term region-wide declines in Caribbean corals. *Science* **301**, 958–960 (2003).
- Williams, S. T., Knowlton, N., Weight, L. A. & Jara, J. A. Evidence for three major clades within the snapping shrimp genus *Alpheus* inferred from nuclear and mitochondrial gene sequence data. *Mol. Phylog. Evol.* **20**, 375–389 (2001).
- Lopez, J. V. & Knowlton, N. Discrimination of sibling species in the *Montastraea annularis* complex using multiple genetic loci. *Proc. 8th Int. Coral Reef Symp.* **2**, 1613–1618 (1997).
- Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods)*, Version 4.0b10 (Sinauer, Sunderland, MA, 2002).
- Kimura, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
- Posada, D. & Crandall, K. A. Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Kimura, M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl Acad. Sci. USA* **78**, 454–458 (1981).

- Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Jara, E. Gomez, M. Hatta and staff of the Palau International Coral Reef Center for their assistance in the field and laboratory, and J. Jackson and R. Grosberg for comments on the manuscript. Financial support came from the National Science Foundation, the Smithsonian Institution, the Scripps Institution of Oceanography, and the Conselho Nacional de Pesquisas (CNPq).

Authors' contributions H.F. performed the collections and genetic analyses, A.F.B. the morphological and palaeontological interpretations, C.A.C. the coral molecular systematics and collections in Taiwan, G.P. the coral systematics and collections in Palau, A.S.-C. the molecular systematics and collections in Brazil, K.I. the collections in Okinawa, and N.K. the coral systematics, financial/logistic support and manuscript preparation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.K. (nknowlton@ucsd.edu). DNA sequences are available in DDBJ (accession numbers AB117222–AB117388 and AB118246–AB118435).

Phylogenetic constraints and adaptation explain food-web structure

Marie-France Cattin¹, Louis-Félix Bersier^{1,2}, Carolin Banašek-Richter¹, Richard Baltensperger³ & Jean-Pierre Gabriel³

¹Institut de Zoologie, University of Neuchâtel, Rue Emile-Argand 11, CP2, CH-2007 Neuchâtel, Switzerland

²Chair of Statistics, Department of Mathematics, Swiss Federal Institute of Technology, CH 1015 Lausanne, Switzerland

³Department of Mathematics, Université de Fribourg, Pérolles, CH 1700 Fribourg, Switzerland

Food webs are descriptions of who eats whom in an ecosystem. Although extremely complex and variable, their structure possesses basic regularities^{1–6}. A fascinating question is to find a simple model capturing the underlying processes behind these repeatable patterns. Until now, two models have been devised for the description of trophic interactions within a natural community^{7,8}. Both are essentially based on the concept of ecological niche, with the consumers organized along a single niche dimension; for example, prey size^{8,9}. Unfortunately, they fail to describe adequately recent and high-quality data. Here, we propose a new model built on the hypothesis that any species' diet is the consequence of phylogenetic constraints and adaptation. Simple rules incorporating both concepts yield food webs whose structure is very close to real data. Consumers are organized in groups forming a nested hierarchy, which better reflects the complexity and multidimensionality of most natural systems.

A central issue in ecology is to uncover the basic determinants of the distribution of trophic interactions among the members of natural communities⁹. The architecture of interactions affects the stability properties of dynamical models of food webs^{2,10,11}. Therefore, a full understanding of dynamic ecosystems cannot be achieved at the economy of assuming a static structure of food webs, as was the case in the pioneering works on stability and complexity that considered interactions to be random^{12,13}. It has been shown unambiguously that real food webs are different from randomly connected networks, and that such a null-model cannot account for the observed properties of the highest-quality food webs available^{6,8}. The structural models of trophic interactions proposed

so far are the cascade model⁷ and the niche model⁸. The former assumes that species are ranked from 1 to S (total number of species), and that a consumer preys only on species of lower rank, with probability $P = 2SC/(S - 1)$ (directed connectance $C = L/S^2$, where L is the total number of trophic links). The niche model orders the species according to a randomly drawn 'niche value', n_i ($0 \leq n_i \leq 1$). A consumer eats all species falling into a range whose centre c_i is randomly chosen, with $c_i < n_i$. This restricts diets to being contiguous (Fig. 1a). Contiguity reflects the ecological assumption that diets can be arranged along one niche dimension.

Whereas the cascade model fails to describe real data, the niche model closely predicts properties of recent high-quality food webs⁸. However, it does not totally account for the observed patterns. The model is known to produce only interval food webs (Fig. 2a, b). These interval webs possess the intriguing characteristic that the feeding relationships between consumers can be represented in one dimension^{1,14}, thought to correspond to a body size hierarchy⁹. But, non-interval food webs characterize recent high-quality data⁸. A

more fundamental problem is that a key assumption of the niche model, contiguity of diets, is not observed in real food webs. By constraining consumers to eat all species in a range, the niche model entails no gap in diets for a suitable ordering of the prey (no gap in columns of Fig. 1a). In a food-web matrix, it is often computationally impracticable to find the prey sequence that would yield the minimum number of gaps. However, an irreducible gap can occur with at least three non-monophagous consumers (a triplet; see Fig. 2c). Consequently, we propose a new property of food-web structure—the level of diet discontinuity, D_{diet} —defined as the proportion of triplets with an irreducible gap over the number of possible triplets. The niche model predicts D_{diet} to equal 0. In seven of the largest and highest-quality empirical food webs (see Methods), the mean of D_{diet} is 0.18, a figure significantly different from 0 ($P = 0.006$, one-tailed, one-sample t -test; Table 1).

These difficulties prompted us to take an evolutionary perspective of the basic ecological determinants of trophic interactions. Typically, a consumer's diet is constrained by its phylogenetic origin¹⁵. For example, all warblers of the *Phylloscopus* genus possess a beak suited to preying on insects; all locusts of the Acrididae family have mouthparts and an alimentary tract suited for plant material. To substantiate the connection between phylogenetic origin and diet, we compared the matrices of trophic and taxonomic similarity for five empirical food webs by Mantel tests¹⁶ (see Methods). There was a strong relationship (all P -values < 0.001), indicating that the distribution of trophic interactions is indeed related to the structure imposed by the phylogeny of the species forming the community.

However, the variability seen in the trophic relations between species cannot be explained fully by phylogenetic constraints. Species have to adapt to varying environments in order to survive, diverging from close relatives in their behaviour, and possibly innovating by using new food sources. Within this framework for example, the diet of New Guinea's fruit pigeons¹⁷ may be seen as constrained by the type of beak and digestive system characteristic of Columbiformes, and the partitioning of fruit size by adaptation to the biotic environment. Dietary shifts also occur frequently, especially in higher animals that are known to be opportunistic. In accordance with the macroevolutionary theory¹⁸, we feel that both phylogenetic constraints and adaptation are the essential determi-

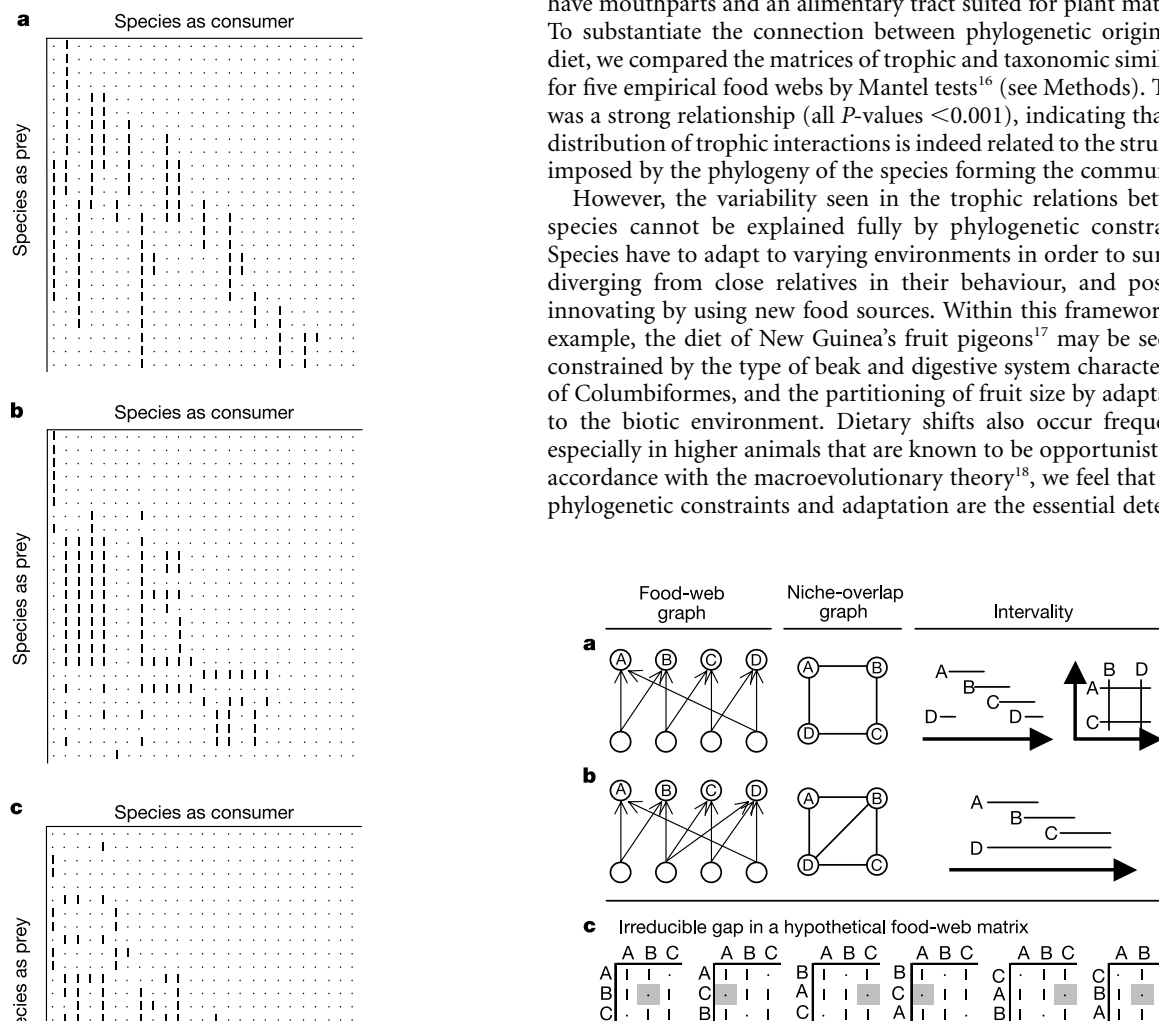


Figure 1 Comparison of one simulation of the niche model (a) and of the nested hierarchy model (c), with respect to a real food web (b); Bridge Brook Lake⁴. The food-web matrix depicts the trophic relationships, where a vertical mark indicates that the species in the column consumes the species in the row (columns and rows contain the same species).

Figure 2 Hypothetical food webs illustrating chordless cycles and intervality, and irreducible gaps. **a, b**, Food-web graph. Circles are species and arrows are flows; A–D are consumers. The niche-overlap graph is an undirected graph where consumers sharing prey are connected. Under 'intervality', graphs depict consumers as segments; segments overlap if consumers share prey. **a**, A chordless and non-interval food web. The niche-overlap graph is chordless because no edge links A to C and/or B to D; the food web is non-interval (consumer D (broken line) cannot be placed in one dimension, because it overlaps with A and C, but not with B; thus, two dimensions are needed). **b**, A non-chordless and interval food web. **c**, Irreducible gaps.

Table 1 Comparison of properties for seven food webs

Property	Skipwith Pond			Little Rock Lake			Bridge Brook Lake			Chesapeake Bay			Ythan Estuary			Coachella Desert			St Martin Island		
<i>T</i>	0.06	[0.04]	0.08	0.05	[0]	0.03	0.10	[0]	0.10	0.14	[0.28]	0.15	0.08	[0.38]	0.07	0.06	[0]	0.08	0.09	[0.17]	0.08
<i>I</i>	0.85	[0.92]	0.82	0.85	[0.87]	0.88	0.73	[0.68]	0.74	0.60	[0.62]	0.59	0.74	[0.53]	0.77	0.86	[0.90]	0.83	0.75	[0.69]	0.77
<i>B</i>	0.09	[0.04]	0.10	0.10	[0.13]	0.08	0.17	[0.32]	0.16	0.26	[0.10]	0.26	0.18	[0.09]	0.17	0.09	[0.10]	0.08	0.17	[0.14]	0.15
<i>Gen_{sd}</i>	0.76	[0.92]	0.83	1.06	[1.40]	1.08	0.98	[1.09]	1.03	1.12	[0.71]	1.17	1.17	[1.16]	1.18	0.81	[0.73]	0.83	1.06	[1.02]	1.09
<i>Vul_{sd}</i>	0.55	[0.54]	0.59	0.59	[0.57]	0.59	0.63	[0.61]	0.72	0.73	[1.03]	0.80	0.66	[1.40]	0.71	0.53	[0.60]	0.60	0.64	[0.78]	0.71
<i>M_{sim}</i>	0.73	[0.76]	0.67	0.69	[0.76]	0.54	0.61	[0.71]	0.56	0.48	[0.34]	0.48	0.58	[0.50]	0.50	0.74	[0.72]	0.67	0.62	[0.54]	0.53
<i>Ch_{mean}</i>	6.23	[4.81]	5.38				4.86	[2.55]	4.30	3.28	[2.77]	3.30	6.39	[4.89]	6.94	6.95	[7.18]	5.78	5.76	[4.20]	5.62
<i>Ch_{sd}</i>	1.56	[1.32]	1.58				1.51	[0.76]	1.52	1.36	[1.14]	1.42	1.89	[1.50]	2.17	1.68	[1.89]	1.64	1.65	[1.30]	1.82
<i>Ch_{log}</i>	4.21	[3.52]	3.70				3.20	[2.62]	2.90	2.35	[2.11]	2.32	4.39	[3.99]	4.38	4.62	[4.90]	4.04	3.96	[3.52]	3.71
<i>Lo</i>	0.25	[0.12]	0.40				0.05	[0]	0.13	0.00	[0.24]	0.02	0.01	[0]	0.06	0.26	[0]	0.42	0.03	[0]	0.11
<i>Can_{sp}</i>	0.43	[0.32]	0.10	0.14	[0.15]	0.03	0.21	[0.12]	0.03	0.08	[0]	0.00	0.07	[0.04]	0.01	0.44	[0.66]	0.09	0.13	[0]	0.02
<i>O</i>	0.77	[0.6]	0.76				0.61	[0.36]	0.60	0.42	[0.38]	0.41	0.61	[0.53]	0.61	0.75	[0.79]	0.78	0.63	[0.60]	0.63
<i>Cy₄</i>	0	[0]	0	0	[639]	626	0	[0]	0	0	[0]	0	0	[206]	18	0	[36]	1	0	[226]	6
<i>D_{diet}</i>	0	[0.35]	0.16	0	[0.08]	0.09	0	[0.004]	0.09	0	[0.08]	0.02	0	[0.32]	0.02	0	[0.28]	0.17	0	[0.17]	0.07

For each food web, empirical values are in brackets, means (medians for *Cy₄*) for the niche model are on the left, and those for the nested hierarchy model on the right. See Methods for a description of the properties of the food webs. Little Rock Lake food web is too large for *Ch_{mean}*, *Ch_{sd}*, *Ch_{log}*, *Lo* and *O* to be computed in a reasonable length of time.

nants of the distribution of trophic interactions in any community.

We devised simple rules to generate food-web matrices according to both concepts. A food-web matrix is an *S*-by-*S* matrix with species as prey in rows and the same list of species as consumers in columns, with a '1' indicating that the species in the column consumes the species in the row (see Fig. 1). As with the cascade and niche models, parameters *S* and *L* are needed. We followed ref. 8 to set the number of links per consumer, which requires species to be ordered according to their 'niche value'; species with the smallest niche value tend to have the smallest number of prey (see Methods). The trophic links are attributed to consumers in a two-stage process, starting with the smallest niche value. In stage one, prey species of consumer *i* is randomly chosen among species with rank $< i$. Depending on this randomly chosen prey *j*, two cases are possible: (1) prey *j* has no consumer and therefore the next prey of consumer *i* will again be randomly attributed (with rank of prey $< i$); (2) prey *j* already has one or more consumers and therefore consumer *i* joins the group of species *j*'s consumers (that is, all consumers sharing at least one prey, with at least one consumer of this group feeding on *j*), and the next prey of consumer *i* is then randomly chosen among the set of prey of this group. However, if the number of prey in the group is too small for choosing all remaining prey of consumer *i*, the remaining prey are again (randomly) chosen among prey without consumers (with rank of prey $< i$). The second stage is needed if prey still cannot be attributed; remaining prey of consumers for which prey could not be attributed in stage 1 are randomly chosen (prey species can have rank $\geq i$).

By creating groups of consumers, stage one (2) expresses the part in food-web organization that is determined by phylogenetic constraints. Links attributed to species free of consumers, and links distributed randomly in the second stage, render the adaptation of consumers to new prey. In forcing consumers to form various trophic groups, our 'nested-hierarchy model' escapes the one-dimensional nature of former models, and better reflects the kind of hierarchies emanating from the phylogenetic structure of the community.

We analysed 12 properties of food webs generated by the niche model as opposed to food webs generated by our model (see Methods). The results of 1,000 simulations of each model for webs with *S* = 20–100 (*S* = 20–50 for computer-intensive properties) and *C* = 0.11 (ref. 19) show that both models yield quite similar results (Fig. 3). When compared with the seven reference food webs, we observe that the niche model performs slightly better for five properties (*Gen_{sd}*, *M_{sim}*, *Lo*, *Can_{sp}*, *O*), that our model is better for four properties (*B*, *Vul_{sd}*, *Ch_{mean}*, *Ch_{log}*), and that both models are equal for *T*, *I* and *Ch_{sd}* (Wilcoxon paired-sample tests, Table 1; see Methods for definitions). In sum, our model performs as well as the niche model—itself outperforming the cascade model by one order of magnitude—in the prediction of these standard food-web descriptors.

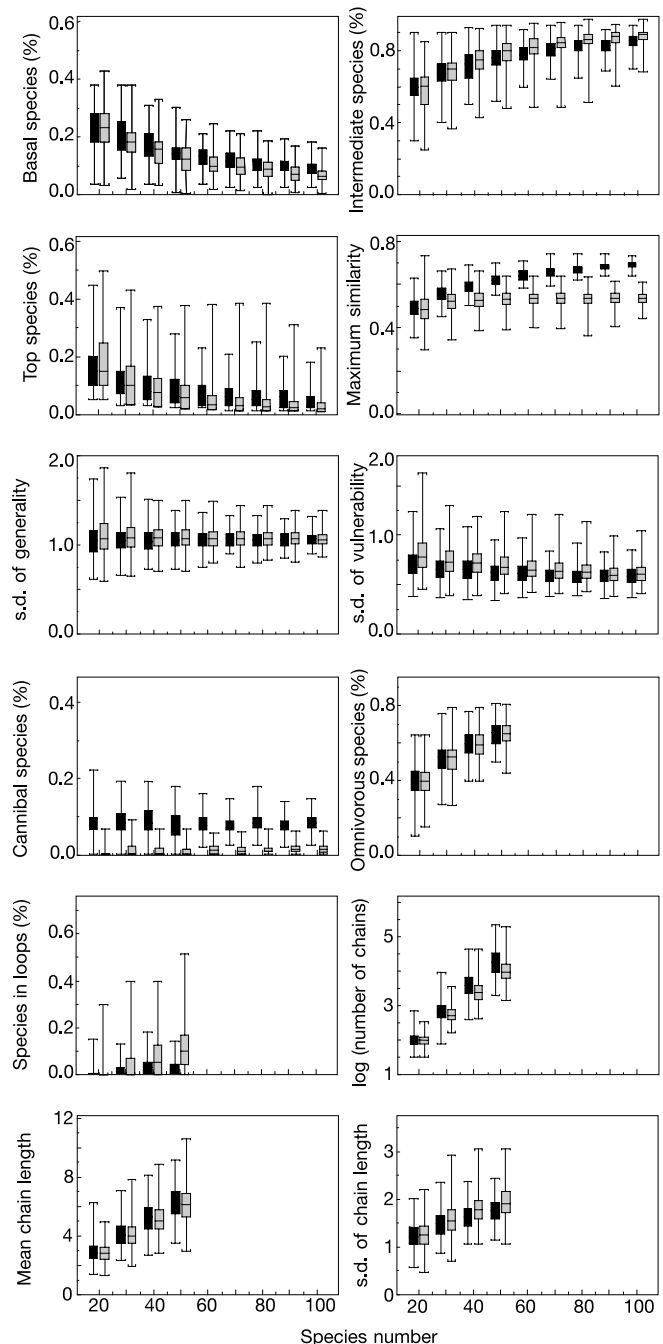


Figure 3 Comparison of 12 properties (see Methods) of food webs generated by the niche model (black boxes) and the nested-hierarchy model (grey boxes).

The nested-hierarchy model surpasses the niche model for two additional properties: D_{diet} and the number of chordless cycles, Cy_4 (Table 1). The latter is closely linked to intervality, as interval food webs have no chordless cycles² (Fig. 2); Cy_4 may thus be thought of as the degree of departure from intervality (that is, from the possibility of ordering the consumers along a single dimension). Finally, the nested-hierarchy model generates food-web matrices whose global structure is close to the observed ones (Fig. 1c).

There are surely other conceivable ways in which to include phylogenetic constraints and adaptation; the one we devised was for us the most intuitive and most parsimonious. The fact that such simple rules are able to faithfully reproduce objects as complex as food webs is a strong hint at their usefulness; for example, as a roadmap for the interactions in dynamical models. What we perceive to be of higher importance than details of model construction are the processes behind the nested-hierarchy model. We have shown how phylogeny is intimately linked to trophic structure in natural communities, and how, once included in a model, it closely predicts observed patterns. When considering trophic links, body size is without doubt another constraint that will limit possible interactions²⁰. Yet body size varies widely within trophic groups, and the size distributions of different trophic groups can overlap extensively—simply think of the herbivores and carnivores in African savannahs. As a consequence, body size is of secondary importance in explaining food-web structure when compared with phylogeny.

Compared with the niche and the cascade models, we impose a sequence in the attribution of links, because a consumer's diet will depend on the species already present. Accordingly, prey are first attributed to primary consumers, as expected in the course of community development. This sequential process is inspired by the niche-hierarchy model²¹, an assembly rule stating that species joining a community will be successful only if they compete within single guilds. For example, it would not be possible for a bird feeding on fruits and insects to enter a community if guilds of frugivores and insectivores are already present. The niche-hierarchy model will also generate nested hierarchies with groups defined by their feeding habits. In our model, we propose that this grouping is constrained by phylogeny, but we permit species to break the strict rule of the niche-hierarchy model by allowing consumers to prey on species outside their group. The niche-hierarchy model has recently been extended to explain patterns of species abundances²². An intriguing perspective would be to combine the niche-hierarchy with our nested-hierarchy model. It will offer a framework to get a sharper understanding of quantitative food webs, including the relationship between abundance, body size and trophic structure²⁰. □

Methods

Data set

The food webs considered here and in ref. 8 are the following (N = total number of taxa, S = total number of trophic species, C = directed connectance): Skipwith pond²³ (N = 35, S = 25, C = 0.32); Chesapeake Bay²⁴ (N = 33, S = 31, C = 0.072); Ythan Estuary²⁵ (N = 92, S = 79, C = 0.061); Coachella Valley²⁶ (N = 30, S = 29, C = 0.31); St Martin Island²⁷ (N = 44, S = 42, C = 0.12); Little Rock Lake²⁸ (N = 180, S = 94, C = 0.12); and Bridge Brook Lake⁸ (N = 75, S = 25, C = 0.17). A trophic species is defined as a set of taxa sharing the same prey and consumers⁸.

Mantel test

Trophic similarity between two species i and j was measured with the index of Jaccard²⁹; that is, as the number of prey and of consumers shared by i and j divided by the pair's total number of prey and consumers. To measure taxonomic similarity, we first ascribed to each taxon (for economy, we use the word 'species' in the text, but prey and consumers are often described at a coarser taxonomic level) its taxonomic membership. We used ten levels: (1) kingdom, (2) superphylum, (3) phylum, (4) subphylum, (5) class, (6) subclass/superorder, (7) order, (8) suborder/superfamily, (9) family, (10) genus. Taxonomic similarity between two taxa, i and j , was measured as the value of the most precise common taxonomic level (for example, ten for two species of the same genus) divided by one plus the value of the most detailed level of any of both taxa. We performed a Mantel test¹⁶ to compare the matrices of trophic and taxonomic similarity of Skipwith pond, Chesapeake Bay, Ythan Estuary, Coachella Valley and St Martin Island. Real taxa and not trophic

species were used. Little Rock Lake and Bridge Brook Lake were not considered because taxonomic information was used to infer trophic interactions.

Model

To determine the number of links per consumer l_i , we assigned each species a niche value n_i uniformly drawn from $[0,1]$ and sorted species according to their n_i in ascending order. Each l_i was obtained by multiplying n_i by a value drawn from $[0,1]$, using a β -distribution with expected value $2C$ and with $\alpha = 1$ (see ref. 8). To obtain the desired L , each consumer's l_i was divided by the sum of l_i over all species and multiplied by L . If the resulting l_i exceeded $S - 1$, we arbitrarily reduced it to $S - 1$. To ensure at least one basal species, the species with the smallest n_i had no prey. For computational reasons, we imposed the presence of at least one top species (the species with the largest n_i has no consumer). We excluded webs with connectance less than 97% of the desired C , with disconnected species, or with species sharing the same consumers and prey.

Properties

The following 14 properties were calculated for each web: B , I , and T , the proportions of basal (without prey), intermediate (with both predators and prey) and top (without predators) species, respectively; Gen_{sd} and Vul_{sd} , the standard deviations (s.d.) of generality and vulnerability⁸, which measure the variation in prey and predator numbers per species, respectively; M_{sim} , the mean maximum similarity of a web, which is the average of all species' largest value for the trophic similarity; Can_{sp} , the proportion of cannibalistic species; O , the proportion of omnivores (that is, species that consume at least two species and have food chains of different lengths); Ch_{log} , the log of the number of food chains (a food chain is a linked path between any species to a basal species); Ch_{mean} and Ch_{sd} , the mean and the s.d. of food chain lengths; Lo , the proportion of species involved in loops other than cannibalistic loops (that is, parts of food chains that include the same species twice). Links involved in loops are ignored for the computation of Ch_{log} , Ch_{mean} , Ch_{sd} and O . The final two properties are Cy_4 , the number of chordless cycles of length four (see Fig. 2), and D_{diet} , the level of diet discontinuity, which measures the proportion of triplets of consumers whose prey cannot be ordered so that the three diets are fully contiguous.

Received 27 October 2003; accepted 12 January 2004; doi:10.1038/nature02327.

- Cohen, J. E. *Food Webs and Niche Space* (Princeton Univ. Press, Princeton, New Jersey, 1978).
- Sugihara, G. *Niche Hierarchy: Structure Assembly and Organization in Natural Communities*. PhD Thesis, Princeton Univ., Princeton (1982).
- Cohen, J. E., Briand, F. & Newman, C. N. *Community Food Webs: Data and Theory* (Springer, Berlin, 1990).
- Havens, K. E. Scale and Structure in Natural Food Webs. *Science* **257**, 1107–1109 (1992).
- Bersier, L. F. & Sugihara, G. Scaling regions for food web properties. *Proc. Natl Acad. Sci. USA* **94**, 1247–1251 (1997).
- Solow, A. R. & Beet, A. R. On lumping species in food webs. *Ecology* **79**, 2013–2018 (1998).
- Cohen, J. E. & Newman, C. M. A stochastic theory of community food webs I. Models and aggregated data. *Proc. R. Soc. Lond. B* **224**, 449–461 (1985).
- Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
- Lawton, J. H. & Warren, P. H. Static and dynamic explanations for patterns in food webs. *Trends Ecol. Evol.* **3**, 242–245 (1988).
- DeAngelis, D. L. Stability and connectance in food web models. *Ecology* **56**, 238–243 (1975).
- Kondoh, M. Foraging adaptation and the relationship between food-web complexity and stability. *Science* **299**, 1388–1391 (2003).
- Gardner, M. & Ashby, W. Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature* **228**, 784 (1970).
- May, R. M. *Stability and Complexity of Model Ecosystems* (Princeton Univ. Press, Princeton, New Jersey, 1974).
- Pimm, S. L. *Food Webs* (Chapman and Hall, London, 1982).
- Cousins, S. H. The trophic continuum in marine ecosystems: Structure and equations for a predictive model. *Can. J. Fish. Aquat. Sci.* **213**, 76–93 (1985).
- Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier, Amsterdam, 1998).
- Diamond, J. M. in *Ecology and Evolution of Communities* (eds Cody, M. L. & Diamond, J. M.) 342–444 (Belknap/Harvard Univ. Press, Cambridge, Massachusetts, 1975).
- Price, P. W. *Macroevolutionary Theory on Macroecological Patterns* (Cambridge Univ. Press, Cambridge, 2003).
- Martinez, N. D. Constant connectance in community food webs. *Am. Nat.* **139**, 1208–1218 (1992).
- Cohen, J. E., Jonsson, T. & Carpenter, S. R. Ecological community description using the food web, species abundance, and body size. *Proc. Natl Acad. Sci. USA* **100**, 1781–1786 (2003).
- Sugihara, G. in *Population Biology. Proceedings of Symposia in Applied Mathematics* (ed. Levin, S. A.) 83–101 (American Mathematical Society, Providence, Rhode Island, 1984).
- Sugihara, G., Bersier, L. F., Southwood, T. R. E., Pimm, S. L. & May, R. M. Predicted correspondence between species abundances and dendrograms of niche similarities. *Proc. Natl Acad. Sci. USA* **100**, 5246–5251 (2003).
- Warren, P. H. Spatial and temporal variation in the structure of a freshwater food web. *Oikos* **55**, 299–311 (1989).
- Baird, D. & Ulanowicz, R. E. The seasonal dynamics of the Chesapeake bay ecosystem. *Ecol. Monogr.* **59**, 329–364 (1989).
- Hall, S. J. & Raffaelli, D. Food-web patterns: lessons from a species-rich web. *J. Anim. Ecol.* **60**, 823–842 (1991).
- Polis, G. A. Complex trophic interactions in deserts: an empirical critique of food-web theory. *Am. Nat.* **138**, 123–155 (1991).
- Goldwasser, L. & Roughgarden, J. Construction and analysis of a large Caribbean food web. *Ecology* **74**, 1216–1233 (1993).
- Martinez, N. D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).

29. Jaccard, P. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaudoise Sci. Nat.* **44**, 223–270 (1908).

Acknowledgements We thank J. Bascompte, S. Cousins, S. Hubbell, R. Naisbit and P. Warren for useful comments. This work was funded by the Swiss National Science Foundation, the Novartis Foundation, and partly by the National Center of Competence in Research 'Plant Survival'.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.-F.B. (louis-felix.bersier@unine.ch).

Global organization of metabolic fluxes in the bacterium *Escherichia coli*

E. Almaas¹, B. Kovács^{1,2}, T. Vicsek², Z. N. Oltvai³ & A.-L. Barabási¹

¹Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

²Biological Physics Department and Research Group of HAS, Eötvös University, H-1117 Budapest, Hungary

³Department of Pathology, Northwestern University, Chicago, Illinois 60611, USA

Cellular metabolism, the integrated interconversion of thousands of metabolic substrates through enzyme-catalysed biochemical reactions, is the most investigated complex intracellular web of molecular interactions. Although the topological organization of individual reactions into metabolic networks is well understood^{1–4}, the principles that govern their global functional use under different growth conditions raise many unanswered questions^{5–7}. By implementing a flux balance analysis^{8–12} of the metabolism of *Escherichia coli* strain MG1655, here we show that network use is highly uneven. Whereas most metabolic reactions have low fluxes, the overall activity of the metabolism is dominated by several reactions with very high fluxes. *E. coli* responds to changes in growth conditions by reorganizing the rates of selected fluxes predominantly within this high-flux backbone. This behaviour probably represents a universal feature of metabolic activity in all cells, with potential implications for metabolic engineering.

To identify the interplay between the underlying topology^{1–3} of the metabolic network of *E. coli* K12 MG1655 and its functional organization, we focus on the global features of potentially achievable flux states in this model organism with a fully sequenced and annotated genome^{13,14}. In accordance with flux balance analysis (FBA)^{8–12}, we first identified the solution space (that is, all possible flux states under a given condition) by using constraints imposed by the conservation of mass and the stoichiometry of the reaction system for the reconstructed *E. coli* metabolic network^{8–12}. Assuming that cellular metabolism is in a steady state and optimized for the maximal growth rate, FBA allows us to calculate the flux for each reaction using linear optimization^{8–11}, which provides a measure of the relative activity of each reaction.

As previously shown^{8–10}, the steady state and optimality approximations offer experimentally verifiable predictions on the flux states of the cell. Under any condition, however, there are also expected differences, some stemming from the fact that there are transient effects and that the growth rate is not always exactly optimal¹². A marked feature of the obtained flux distribution is its overall inhomogeneity: reactions with fluxes spanning several orders of

magnitude coexist under the same conditions. For example, under optimal growth conditions in a glutamate-rich culture, the dimensionless flux of the succinyl coenzyme A synthetase reaction is 0.185, whereas the flux of the aspartate oxidase reaction is four orders of magnitude smaller, with a value of 2.2×10^{-5} in dimensionless units (the flux vector is normalized to unity).

To characterize the coexistence of such widely different flux values, we plot the flux distribution for active (non-zero flux) reactions of *E. coli* grown in a glutamate- or succinate-rich substrate (Fig. 1a). The distribution is best-fitted with a power law with a small flux constant, indicating that the probability that a reaction has flux ν follows $P(\nu) \propto (\nu + \nu_0)^{-\alpha}$, where the constant is $\nu_0 = 0.0003$ and the flux exponent has the value $\alpha = 1.5$. The observed power law is also consistent with published experimental data. Indeed, the flux distribution obtained from the measured fluxes of the central metabolism of *E. coli*¹⁵ is best-fitted with the power-law form (Fig. 1d). As the central metabolism is characterized by high fluxes, the small-flux saturation seen in Fig. 1a is absent from these data.

To examine whether the observed flux distribution is independent of the environmental conditions, we mimicked the influence of various growth conditions by randomly choosing 10%, 50% or 80% of the 96 potential substrates that *E. coli* can consume (the input substrates are listed in Supplementary Table S2). After optimizing the growth rate, we find that the power-law distribution of metabolic fluxes is independent of the external conditions (Fig. 1b). As the metabolic activity of *E. coli* frequently deviates from the optimal growth state under variable growth conditions^{10,12}, we inspected whether the wide flux distribution is also present in non-optimal

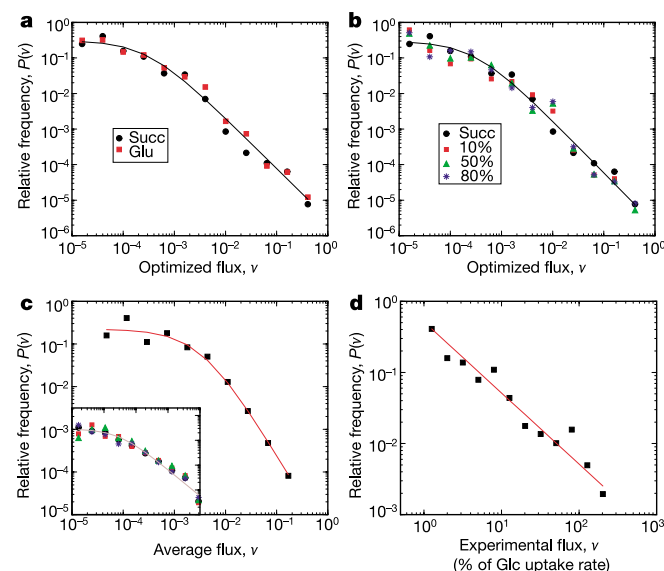


Figure 1 Characterizing the overall flux organization of the *E. coli* metabolic network. **a**, Flux distribution for optimized biomass production on succinate (black) and glutamate (red) substrates. Solid line corresponds to the power-law fit $P(\nu) \propto (\nu + \nu_0)^{-\alpha}$, with $\nu_0 = 0.0003$ and $\alpha = 1.5$. **b**, Flux distribution for optimized biomass on succinate (black) substrate with an additional 10% (red), 50% (green) and 80% (blue) randomly chosen subsets of the 96 input channels turned on. The flux distribution was averaged over 5,000 independent random choices of uptake metabolites. The resulting flux distribution can be fitted (solid line) by a power law with parameters $\nu_0 = 0.0004$ and $\alpha = 1.5$. **c**, Flux distribution from the non-optimized hit-and-run sampling method^{16,17} of the *E. coli* solution space. The solid line is the best fit, with $\nu_0 = 0.003$ and $\alpha = 2$. Inset shows the flux distribution in four randomly chosen sample points. **d**, The distribution of experimentally determined fluxes (see ref. 15) from the central metabolism of *E. coli* shows power-law behaviour, with a best fit to $P(\nu) \propto \nu^{-\alpha}$ with $\alpha = 1$.

states. For this, we implemented a “hit-and-run” method^{16,17} that randomly samples the full solution space, enabling us to calculate the flux for each reaction in 50,000 distinct non-optimal states.

Although the obtained average flux distribution is consistent in shape and flux ranges with those obtained by the optimal FBA, the flux exponent is slightly larger ($\alpha = 2$, Fig. 1c), and the quality of the scaling is slightly weaker. Notably, many individual non-optimal states (Fig. 1c, inset) are consistent with an exponent $\alpha = 1$, in accord with the experimental results (Fig. 1d) and supporting the prediction^{10,12} that these organisms may not have achieved optimality. These findings imply that the observed flux distribution is a generic feature of flux conservation¹⁸ on a scale-free network¹⁹, as it is independent of the optimal or non-optimal nature of the growth rate or the growth conditions. The exponent, however, may depend on the organism's position in the solution space.

An exponentially decaying distribution (for example, a gaussian) would predict that under a given condition most reactions are characterized by comparable fluxes. By contrast, the identified power-law flux distribution suggests a highly uneven use: most reactions have small fluxes and coexist with a few reactions with very high flux values. Hence, the biochemical activity of the metabolism is dominated by several ‘hot’ reactions, which are embedded in a network of mostly small-flux reactions. The fact that the observed flux exponent α is less or equal to 2 implies that the first and higher moments of the flux distribution are not defined from a mathematical perspective (they diverge for an infinite system). The divergence of the average flux $\langle \nu \rangle$ (first moment) indicates that no

flux value can be designated as characteristic for the system. Although $\langle \nu \rangle$ can be calculated numerically by averaging over the flux of all reactions, with $\langle \nu_g \rangle = 0.0171$ and $\langle \nu_s \rangle = 0.0173$ for glutamate and succinate uptakes, respectively, these values are by no means indicative of the overall activity of the metabolic reactions. Indeed, most reactions (86% for glutamate and 89% for succinate uptake) have a smaller flux than this average, and a few reactions have an activity that is orders of magnitude higher.

The observed flux distribution is compatible with two different potential local flux structures. A homogeneous local organization would imply that all reactions producing (consuming) a given metabolite have comparable fluxes; however, a more delocalized ‘hot backbone’ is expected if the local flux organization is heterogeneous, such that each metabolite has a dominant source (consuming) reaction (Supplementary Fig. S7). To distinguish between these two schemes for each metabolite i produced (consumed) by k reactions, we define the measure²⁰

$$Y(k, i) = \sum_{j=1}^k \left(\frac{\hat{\nu}_{ij}}{\sum_{l=1}^k \hat{\nu}_{il}} \right) \quad (1)$$

where $\hat{\nu}_{ij}$ is the mass carried by reaction j which produces (consumes) metabolite i . If all reactions producing (consuming) metabolite i have comparable $\hat{\nu}_{ij}$ values, $Y(k, i)$ scales as $1/k$. If, however, the activity of a single reaction dominates equation (1), we expect $Y(k, i) \propto 1$, in other words, $Y(k, i)$ is independent of k . For the *E. coli* metabolism optimized for succinate and glutamate uptake, we find

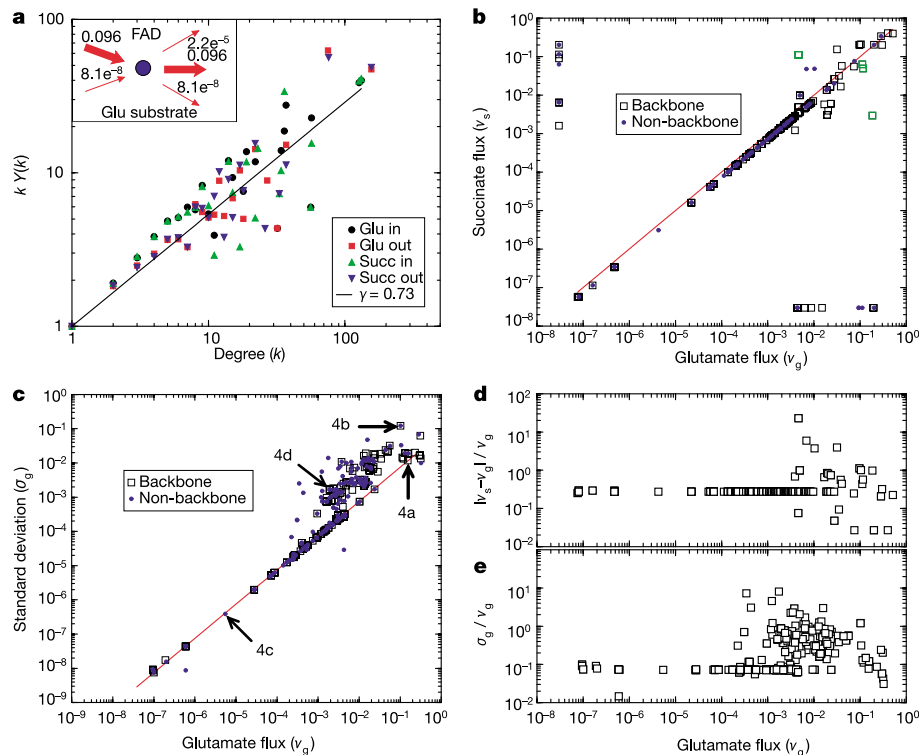


Figure 2 Characterizing the local inhomogeneity of the metabolic flux distribution. **a**, Measured $kY(k)$ shown as a function of k for incoming and outgoing reactions, averaged over all metabolites, indicates that $Y(k) \propto k^{-0.27}$, as the solid line has the slope $\gamma = 0.73$. Inset shows non-zero mass flows, $\hat{\nu}_{ij}$, producing (consuming) FAD on a glutamate-rich substrate. **b**, Change in the flux of individual reactions when switching from glutamate-rich (horizontal axis, ν_g) to succinate-rich (vertical axis, ν_s) conditions. Reactions with negligible flux changes follow the diagonal (solid line). Some reactions are turned off in only one of the conditions (shown close to the coordinate axes). Reactions belonging to the HFB (see Fig. 3) are indicated by black squares, the rest are indicated by

blue dots. Reactions in which the direction of the flux is reversed are coloured green. **c**, Absolute value of glutamate flux (ν_g) for 50% randomly chosen input channels (averaged over 5,000 realizations) plotted against the standard deviation of the same reaction. The red line, corresponding to $y = 0.075x$, is shown for reference. Numbers 4a–4d identify the reactions whose distribution is shown in Fig. 4. **d**, Relative flux change, $\Delta \nu = (|\nu_s - \nu_g| / \nu_g)$, as a function of ν_g for the conditions shown in **b**. **e**, Relative flux fluctuation, σ_g / ν_g , per reaction for the conditions shown in **c**, again emphasizing that large changes are limited to high-flux reactions.

that both the 'in' and 'out' degrees follow $Y(k,i) \propto k^{-0.27}$ (Fig. 2a), representing an intermediate behaviour between the two extreme cases. This indicates that the large-scale inhomogeneity observed in the overall flux distribution is also increasingly valid at the level of the individual metabolites: the more reactions that consume (produce) a given metabolite, the more likely it is that a single reaction carries most of the flux.

Such inhomogeneity is obvious, for example, for flavin adenine dinucleotide (FAD), whose production (consumption) is dominated by only one of the two (three) contributing reactions (Fig. 2a, inset), and the high-flux reactions are catalysed by succinate dehydrogenase complex, EC 1.3.5.1 (succinate dehydrogenase, EC 1.3.99.1). We find that $Y(k,i)$ scales in a similar fashion when *E. coli* is grown in rich Luria-Bertani medium or non-optimized configurations, indicating that the local inhomogeneity is not a unique feature of the optimized state or a specific growth condition but represents a generic property of the local flux distribution (see Supplementary Information).

The local flux inhomogeneity indicates that for most metabolites we can identify a single reaction that dominates its production (consumption). This observation can be turned into a simple algorithm, which for each metabolite systematically removes all reactions but the one providing the largest incoming (outgoing) flux contribution. The algorithm uncovers the 'high-flux backbone' (HFB) of the metabolism, a distinct structure of linked reactions that forms a giant component²¹ with a star-like topology (Fig. 3 and Supplementary Fig. S12), including most metabolites produced under the given growth condition. Only a few pathways appear disconnected, indicating that although these pathways are part of the HFB, their end product is only the second-most important source for another HFB metabolite. Of note, groups of individual HFB reactions largely overlap with the traditional, biochemistry-based partitioning of cellular metabolism: for example, all metabolites of the citric acid cycle of *E. coli* are recovered, and so are a considerable fraction of other important pathways, such as those involved in histidine, murein and purine biosynthesis. The HFB

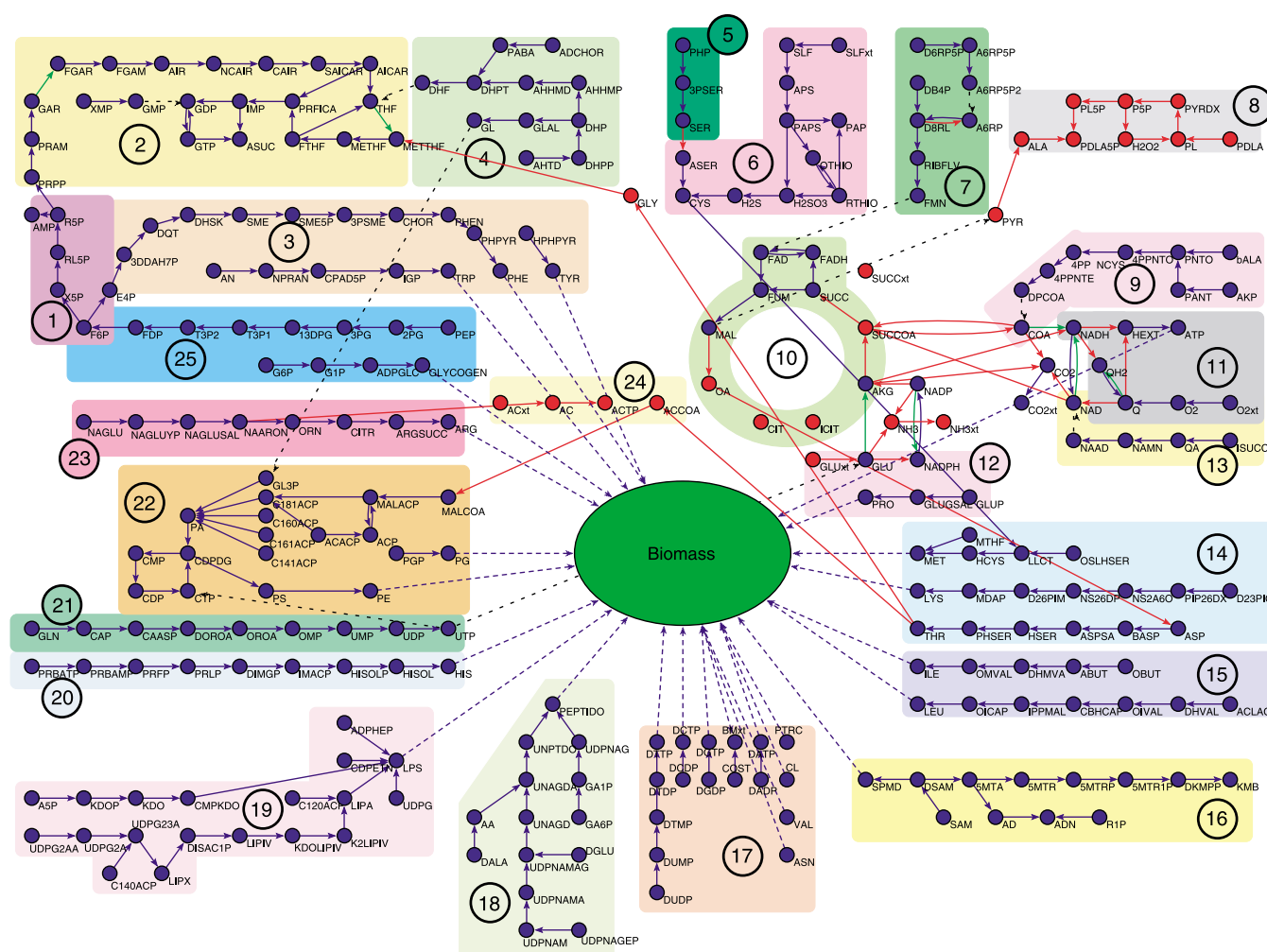


Figure 3 High-flux backbone for FBA-optimized metabolic network of *E. coli* on a glutamate-rich substrate (see Supplementary Fig. S12b for succinate-rich substrate). For any two metabolites (called for example, A and B) are connected with a directed link pointing from A to B only if the reaction with maximal flux consuming A is the reaction with maximal flux producing B. Shown are all metabolites that have at least one neighbour after completing this procedure. The background colours denote different known biochemical pathways. Metabolites (vertices) coloured blue have at least one neighbour in common in glutamate- and succinate-rich substrates, and those coloured red have none. Reactions

(lines) are coloured blue if they are identical in glutamate- and succinate-rich substrates, green if a different reaction connects the same neighbour pair, and red if this is a new neighbour pair. Black dotted lines indicate where the disconnected pathways, for example, folate biosynthesis (4), would connect to the cluster through a link that is not part of the HFB. Thus, the red nodes and links highlight the predicted changes in the HFB when shifting *E. coli* from glutamate- to succinate-rich media. Dashed lines indicate links to the biomass growth reaction. Numbers identifying the various biochemical pathways are listed in the Supplementary Information.

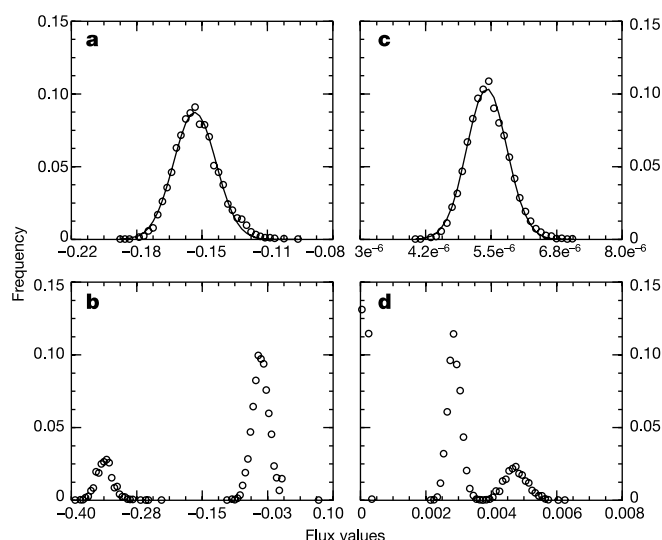


Figure 4 Effect of growth conditions on individual fluxes. Shown is the flux distribution for four select *E. coli* reactions in a 50% random environment (see Fig. 2c). **a**, Triphosphate isomerase; **b**, carbon dioxide transport; **c**, NAD kinase; **d**, guanosine kinase. Reactions on the $\sigma \propto \nu$ curve have gaussian distributions (**a** and **c**); reactions off this curve have multimodal distributions (**b** and **d**), showing several discrete flux values under diverse conditions. Solid curves correspond to gaussians derived using the calculated ν and σ values of -0.15 and 0.012 (**a**) and $5.4e^{-6}$ and $3.9e^{-7}$ (**c**).

captures the subset of reactions that dominate the activity of the metabolism. As such, it offers a complementary approach to elementary flux mode analyses^{22–24}, which successfully capture the available modes of operation for smaller metabolic subnetworks.

As the flux of the individual metabolic reactions depends on the growth conditions, we need to inspect how sensitive the HFB is to changes in the environment. Unexpectedly, Fig. 2b, c, which records the relationship between the individual fluxes under succinate and glutamate uptake, indicates that only the reactions in the high-flux territory undergo noticeable flux changes, whereas the reactions in the intermediate- and low-flux regions remain virtually unaltered (a small shift can be observed, however, because there is a 41% increase in biomass production in glutamate-rich versus succinate-rich media). The observed flux changes correspond to two types of event. First, some pathways are turned off completely (type I reactions), having zero flux under one growth condition and high flux in the other. These reactions are shown as symbols along the horizontal and vertical axis in Fig. 2b. By contrast, other reactions remain active but show an orders-of-magnitude shift in flux under the two different growth conditions (type II reactions). Notably, with two exceptions these marked type II changes are limited to the HFB reactions. The same phenomenon is predicted when we inspect the transition from glucose to succinate uptake, or for transitions among various random uptake conditions (see Supplementary Information).

To test the generality of this finding, we mimicked the effect of various growth conditions by randomly choosing 50% of the potential input substrates, measuring in each input configuration the flux for each reaction. For each reaction, the average flux (ν) and the standard deviation (σ) around this average were determined by averaging over 5,000 random input conditions. For small fluxes, all reactions closely follow a straight line (corresponding to $\sigma \propto \nu$), supporting the above finding that the small fluxes remain essentially unaltered as the external conditions change (Fig. 2c). For the high-flux reactions, however, there are noticeable deviations from this line, indicating that there are considerable variations in flux from

one external condition to the other (Fig. 2d, e). A closer inspection of the flux distribution shows that the reactions on the $\sigma \propto \nu$ curve all have a clear unimodal flux distribution (Fig. 4a, c), indicating that shifts in growth conditions lead to only small changes (within σ) of their flux values. By contrast, the reactions deviating from the $\sigma \propto \nu$ curve have a bi- or trimodal distribution, indicating that under different growth conditions they show several discrete and distinct flux values (Fig. 4b, d).

Therefore, Figs 2c–e, 3 and 4 offer valuable insights into how *E. coli* responds to changes in growth conditions: it activates or deactivates specific metabolic reactions among the HFB metabolites in new ways without altering the identity of the principal pathways that participate in the backbone. For example, switching from a glutamate to a succinate substrate turns off vitamin B6 biosynthesis (type I change), reconnects in other ways several metabolic substrates, such as those of the coenzyme A and NAD biosynthesis pathways as well as the respiration pathways, and modifies the use of the citric acid cycle (Supplementary Fig. S12). Apart from minor changes, the use of the other pathways remains unaltered. These reorganizations result in large, discrete changes in the fluxes of the HFB reactions.

The power-law distribution of metabolic fluxes in the *E. coli* metabolism indicates a highly uneven use of the underlying metabolic network topology. Although wide flux differences among various pathways are known from individual experiments^{15,25–27}, we find that they are part of a scale-invariant continuum, which follows a scaling law. The uneven flux use is present both at the global level (Fig. 1) and at the level of the individual metabolites (Fig. 2a). This observation allows us to uncover automatically the HFB of the metabolism and could provide insights into metabolic organization and regulation as well as offering valuable inputs for metabolic engineering.

The observation and theoretical prediction of a power-law load distribution in simple models (ref. 18 and Supplementary Information), coupled with the presence of a power law in both the optimal and non-optimal flux states, suggests that the metabolic flux organization is a direct consequence of the scale-free topology of the network¹⁹. As all organisms examined so far are characterized by a scale-free metabolic network topology¹, the observed scaling in the flux distribution is probably not limited to *E. coli*, but characterizes all organisms from eukaryotes to archaea. As FBA is available for an increasing number of prokaryotic and eukaryotic organisms, this prediction could be verified both experimentally and theoretically in the near future. Hence, the observed uneven local and global flux distribution seems to be rooted in the subtle, yet generic, interplay of the network's directed topology and flux balance, channelling the numerous small fluxes into high-flux pathways. The dependence of the scaling exponents characterizing the flux distributions on the nature of the optimization process, as well as the experimentally observed exponent, may be a benchmark for future structural and evolutionary models aiming to explain the origin, the organization and the modular structure^{3,4,28,29} of cellular metabolism. □

Methods

Flux balance analysis

Starting from a stoichiometric matrix of the MG1655 strain^{8,9} of *E. coli*, containing 537 metabolites and 739 reactions, the steady-state concentrations of all the metabolites satisfy:

$$\frac{d}{dt}[A_i] = \sum_j S_{ij} \nu_j = 0 \quad (2)$$

where S_{ij} is the stoichiometric coefficient of metabolite A_i in reaction j and ν_j is the flux of reaction j . We use the convention that if metabolite A_i is a substrate (product) in reaction j , then $S_{ij} < 0$ ($S_{ij} > 0$), and we constrain all fluxes to be positive by dividing each reversible reaction into two 'forward' reactions with positive fluxes. Any vector of positive fluxes $\{\nu_j\}$ that satisfies equation (2) corresponds to a state of the metabolic network and, hence, a potential state of operation of the cell. We restrict our study to the subspace of solutions for which all components of ν satisfy the constraint $\nu_j > 0$ (ref. 8). We denote the mass carried

by reaction j producing (consuming) metabolite i by:

$$\dot{v}_{ij} = |S_{ij}|v_j$$

Random uptake conditions

We choose randomly $X\%$ (where $X = 10, 50$ or 80) of the 89 potential input substrates that *E. coli* consumes in addition to the minimal uptake basis. For each of the transport reactions, we set the uptake rate to 20 mmol per gram of dry weight per hour. As there are a very large number of possible combinations of the selected input substrates, we repeat this process 5,000 times and average over each realization.

The hit-and-run method

We select a set of basis vectors spanning the solution space using singular-value decomposition. Because the reaction fluxes must be positive, the 'bouncer' is constrained to the part of the solution space that intersects the positive orthant. We constrain the bouncer within a hypersphere of radius $R_{\max}$ and outside a hypersphere of radius $R_{\min} < R_{\max}$, where we find that the sampling results are independent of the choices of $R_{\min}$ and $R_{\max}$. Starting from a random initial point inside the positive flux cone in a randomly chosen direction, the bouncer travels deterministically a distance d between sample points. Each sample point, corresponding to a solution vector where the components are the individual fluxes, is normalized by projection onto the unit sphere. After every b th bounce off the internal walls of the flux cone, the direction of the bouncer is randomized.

High-flux backbone

For each metabolite we keep only the reactions with the largest flux that produces and consumes the metabolite. Metabolites that are not produced (consumed) are discounted. Subsequently, a directed link is introduced between two metabolites A and B if (1) A is a substrate of the most active reaction producing B, and (2) B is a product of the maximal reaction consuming A. We consider only metabolites that are connected to at least one other metabolite after steps (1) and (2). For clarity, we remove P_i, PP_i and ADP. Further details and figures are provided in the Supplementary Information.

Received 29 August; accepted 12 December 2003; doi:10.1038/nature02289.

- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
- Wagner, A. & Fell, D. A. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B* **268**, 1803–1810 (2001).
- Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
- Holme, P., Huss, M. & Jeong, H. Subnetwork hierarchies of biochemical pathways. *Bioinformatics* **19**, 532–538 (2003).
- Savageau, M. A. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, MA, 1976).
- Heinrich, R. & Schuster, S. *The Regulation of Cellular Systems* (Chapman & Hall, New York, 1996).
- Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behavior* (Cambridge Univ. Press, Cambridge, UK, 1996).
- Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97**, 5528–5533 (2000).
- Edwards, J. S., Ibarra, R. U. & Palsson, B. O. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnol.* **19**, 125–130 (2001).
- Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189 (2002).
- Edwards, J. S., Ramakrishna, R. & Palsson, B. O. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* **77**, 27–36 (2002).
- Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15112–15117 (2002).
- Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474 (1997).
- Gerdes, S. Y. *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185**, 5673–5684 (2003).
- Emmerling, M. *et al.* Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *J. Bacteriol.* **184**, 152–164 (2002).
- Smith, R. L. Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.* **32**, 1296–1308 (1984).
- Lovász, L. Hit-and-run mixes fast. *Math. Program.* **86**, 443–461 (1999).
- Goh, K. I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* **87**, 278701 (2001).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Barthelemy, M., Gondran, B. & Guichard, E. Spatial structure of the Internet traffic. *Physica A* **319**, 633–642 (2003).
- Ma, H. W. & Zeng, A. P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19**, 1423–1430 (2003).
- Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, P. Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.* **343**, 115–124 (1999).
- Schuster, S., Fell, D. A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnol.* **18**, 326–332 (2000).
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193 (2002).
- Sauer, U. *et al.* Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.* **181**, 6679–6688 (1999).
- Canonaco, F. *et al.* Metabolic flux response to phosphoglucose isomerase knock-out in *Escherichia coli* and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiol. Lett.* **204**, 247–252 (2001).

- Fischer, E. & Sauer, U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur. J. Biochem.* **270**, 880–891 (2003).
- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- Wolf, D. M. & Arkin, A. P. Motifs, modules and games in bacteria. *Curr. Opin. Microbiol.* **6**, 125–134 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Bárász, J. Becker, E. Ravasz, A. Vazquez and S. Wuchty for discussions; and B. Palsson and S. Schuster for comments on the manuscript. Research at Eötvös University was supported by the Hungarian National Research Grant Foundation (OTKA), and work at the University of Notre Dame and at Northwestern University was supported by the US Department of Energy, the NIH and the NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.-L.B. (alb@nd.edu).

Sortilin is essential for proNGF-induced neuronal cell death

Anders Nykjaer^{1,2}, Ramee Lee³, Kenneth K. Teng³, Pernille Jansen^{1,4}, Peder Madsen¹, Morten S. Nielsen¹, Christian Jacobsen¹, Marco Kliehmannel⁵, Elisabeth Schwarz⁵, Thomas E. Willnow^{2,4}, Barbara L. Hempstead³ & Claus M. Petersen¹

¹Department of Medical Biochemistry, Ole Worms Allé 170, Aarhus University, and ²ReceptIcon Aps, Gustav Wieds vej 10, DK-8000 Aarhus C, Denmark
³Weill Medical College of Cornell University, New York, New York 10021, USA
⁴Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany
⁵Institute for Biotechnology, Martin-Luther-Universität, Halle-Wittenberg, 06120 Halle, Germany

Sortilin¹ (~95 kDa) is a member of the recently discovered family of Vps10p-domain receptors^{2,3}, and is expressed in a variety of tissues, notably brain, spinal cord and muscle. It acts as a receptor for neurotensin^{4,5}, but predominates in regions of the nervous system that neither synthesize nor respond to this neuropeptide⁶, suggesting that sortilin has additional roles. Sortilin is expressed during embryogenesis⁷ in areas where nerve growth factor (NGF) and its precursor, proNGF, have well-characterized effects^{8,9}. These neurotrophins can be released by neuronal tissues^{8,9}, and they regulate neuronal development through cell survival and cell death signalling. NGF regulates cell survival and cell death via binding to two different receptors, TrkA and p75^{NTR} (ref. 10). In contrast, proNGF selectively induces apoptosis through p75^{NTR} but not TrkA¹¹. However, not all p75^{NTR}-expressing cells respond to proNGF, suggesting that additional membrane proteins are required for the induction of cell death. Here we report that proNGF creates a signalling complex by simultaneously binding to p75^{NTR} and sortilin. Thus sortilin acts as a co-receptor and molecular switch governing the p75^{NTR}-mediated pro-apoptotic signal induced by proNGF.

Binding of NGF was examined by surface-plasmon resonance (SPR). As demonstrated in Fig. 1a, sortilin bound mature NGF with moderate affinity (dissociation constant (K_d) ~90 nM). In contrast, the affinity of NGF for p75^{NTR} and TrkA was high (K_d 1–2 nM), in accordance with previous studies in cells^{11–13}. As the NGF precursor (proNGF) may escape intracellular processing and be released extracellularly, we next examined binding of proNGF^{9,11,14,15}. Whereas lack of processing reduces the affinity of proNGF for p75^{NTR} and TrkA (K_d ~15–20 nM), it results in a much higher affinity (K_d ~5 nM) for sortilin (Fig. 1a). This is surprising because

proNGF has been reported to interact with cellular p75^{NTR}, but not TrkA, with a higher affinity ($K_d \sim 0.2$ nM) than mature NGF, and to selectively induce p75^{NTR}-dependent apoptosis in neurons, smooth muscle cells and oligodendrocytes^{8,11}. Our data may reflect the participation of a p75^{NTR} co-receptor, and this receptor might be sortilin (see below).

To examine the structural basis of proNGF binding, we produced the pro domain of proNGF as a glutathione *S*-transferase fusion protein (GST-pro) (Fig. 1b). The GST-pro protein bound to sortilin with an affinity very similar to that of proNGF ($K_d \sim 8$ nM), but not to p75^{NTR} or TrkA (Fig. 1a). Additional experiments further demonstrated that binding of proNGF to sortilin was inhibited markedly (>75%) by neurotensin, and was almost abolished in the presence of GST-pro (Fig. 1c). Thus, the pro domain constitutes the structural basis for the high-affinity binding between proNGF and sortilin.

We next assessed binding of proNGF to cellular sortilin. We transfected 293 cells without endogenous sortilin with the receptor constructs indicated in Fig. 2a, and evaluated binding of proNGF at 37 °C. Control cells demonstrated no binding or uptake of ligand, whereas cells expressing wild-type sortilin exhibited significant endocytosis of proNGF (Fig. 2b), but not of mature NGF (Fig. 2c). The observed uptake was hampered strongly in the presence of excess neurotensin or GST-pro (data not shown). Furthermore, transfectants expressing a mutant sortilin protein (sortilin(mut)) that accumulates on the plasma membrane owing to disrupted motifs for endocytosis¹⁶, displayed intense surface labelling with proNGF, but little uptake.

Binding and uptake of proNGF was also investigated in cells transfected with TrkA and p75^{NTR}, and in cells expressing each of the two receptors in combination with sortilin (Fig. 2b). TrkA transfectants exhibited a very modest uptake of proNGF, and on co-transfection with sortilin, endocytosis of proNGF was comparable to that observed in cells expressing sortilin alone. In contrast, uptake of mature NGF was efficient in TrkA-expressing cells and was unaffected by co-transfection with sortilin (Fig. 2c). In p75^{NTR} transfectants, proNGF as well as mature NGF was almost exclusively found on the plasma membrane, indicating a slow or insignificant endocytosis (Fig. 2b, c) consistent with prior observations^{17,18}. However, coexpression of p75^{NTR} with sortilin re-established uptake of proNGF, and coexpression with sortilin(mut), as well as with wild-type sortilin, induced a significant increase in surface-associated ligand, suggesting a synergistic rather than a simple

additive effect of sortilin and p75^{NTR} coexpression.

The findings demonstrate that sortilin exhibits negligible binding of NGF, but also that it conveys a significantly higher capacity for uptake of proNGF than either of the two established receptors (that is, p75^{NTR} and TrkA). Moreover, results in double transfectants suggest that sortilin and p75^{NTR} cooperate to promote proNGF binding.

To characterize the molecular mechanisms underlying the putative cooperativity between p75^{NTR} and sortilin, affinity crosslinking was performed. Crosslinking of ¹²⁵I-labelled proNGF to p75^{NTR} and sortilin double transfectants produced labelled complexes of ~ 110 , ~ 140 and ~ 240 kDa (Fig. 3a, lane 1), but was unproductive in single transfectants, suggesting that coexpression of sortilin and p75^{NTR} is required for efficient binding at subnanomolar concentrations of proNGF (Fig. 3a, lanes 5–6). Furthermore, immunoprecipitation with receptor antisera established that both sortilin and p75^{NTR} were components of the crosslinked adducts (Fig. 3a, lanes 7–8). Similar experiments performed on cells coexpressing p75^{NTR} and sortilin(mut), which has a higher surface expression than wild-type sortilin¹⁶, resulted in a quantitative increase in crosslinked complexes (data not shown). Finally, generation of the crosslinking adducts was markedly reduced in the presence of unlabelled proNGF, neurotensin or GST-pro, implying that sortilin, as well as the NGF pro domain, is critical to complex formation (Fig. 3a, lanes 2–4).

We conclude that the ~ 240 kDa adduct probably represents a heterotrimeric complex comprising proNGF, sortilin and p75^{NTR}, whereas the ~ 110 kDa and ~ 140 kDa species constitute proNGF in association with a single receptor. Expression of both receptors is required for efficient binding of proNGF, and our results support a model in which the pro domain and the 'mature' part of proNGF simultaneously engage sortilin and p75^{NTR}, respectively.

Corresponding experiments established that proNGF does not form stable complexes with TrkA (Fig. 3b). In fact, crosslinking with proNGF using cells expressing all three receptors (TrkA, p75^{NTR} and sortilin) resulted in ~ 110 , ~ 140 and ~ 240 kDa adducts that could be precipitated with anti-sortilin (data not shown) and anti-p75^{NTR} antibodies but not with TrkA-specific antiserum. Thus, proNGF discriminates between TrkA and p75^{NTR} in cells that express both receptors in combination with sortilin.

In accordance with previous reports^{12,13}, crosslinking using mature ¹²⁵I-labelled NGF yielded complexes with both p75^{NTR} (~ 90 and ~ 180 kDa) and TrkA (~ 160 kDa) (Fig. 3c). However,

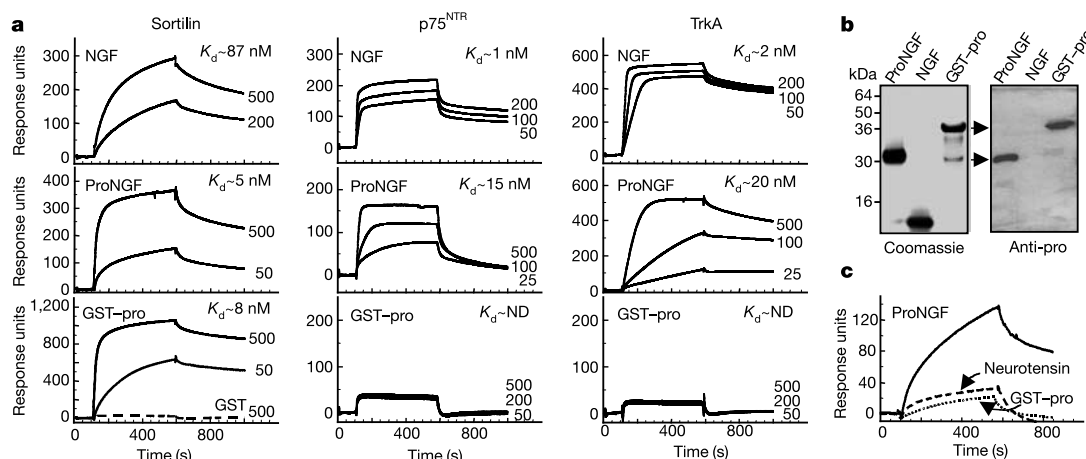


Figure 1 SPR analysis of ligand binding. **a**, Binding of 20–1,000 nM NGF, proNGF and GST-pro to immobilized sortilin (51 fmol mm⁻²), p75^{NTR} (91 fmol mm⁻²) and TrkA (66 fmol mm⁻²). Calculated K_d values are indicated. **b**, SDS-PAGE analysis of ligands used in **a**. A Coomassie-stained gel (left panel) and a western blot (anti-pro domain

antibody; right panel) are shown. **c**, Binding of proNGF (25 nM) to sortilin (66 fmol mm⁻²) in the absence or presence of 10 μ M neurotensin (dashed line) and 5 μ M GST-pro (dotted line). The SPR response obtained for the inhibitors alone has been subtracted.

no additional crosslinked complexes were observed when either of the two was coexpressed with sortilin, and sortilin single transfectants did not bind NGF. These results indicate that sortilin neither interacts with mature NGF nor is part of a complex formed on binding of mature NGF to p75^{NTR} or TrkA.

We next examined whether sortilin and p75^{NTR} physically associate on the cell membrane. Cells expressing both receptors were biolabelled and incubated in the absence or presence of proNGF, followed by treatment with a membrane-impermeable reducible crosslinker, lysis and immunoprecipitation using anti-p75^{NTR} antibodies. Sortilin could be crosslinked directly to p75^{NTR} (Fig. 3d). However, in the presence of proNGF, the relative amount of crosslinked and co-precipitated sortilin increased by about fivefold (3.9% to 18.4%). Equilibrium binding studies were then designed to determine whether p75^{NTR} and sortilin coexpression might influence the specificity and affinity of ligand binding. As demonstrated in Fig. 3e, ¹²⁵I-labelled NGF bound to cells coexpressing sortilin and p75^{NTR} with an estimated K_d of ~1.0 nM. This agrees with previous results¹² obtained in p75^{NTR}-expressing cells, and as sortilin single transfectants did not bind mature NGF (data not shown), the data indicate that mature NGF binds strictly to p75^{NTR}. In contrast, similar experiments with ¹²⁵I-labelled proNGF further indicated that sortilin and p75^{NTR} cooperate in proNGF binding. Thus, cells expressing a single receptor type—either sortilin or p75^{NTR}—did not bind ¹²⁵I-labelled proNGF (data not shown), whereas cells

coexpressing these receptors did. Scatchard analysis (Fig. 3e) suggested fewer binding sites for proNGF than for NGF in the double transfectants, but also a higher affinity for proNGF (K_d ~160 pM) that could not be accounted for by binding to any single receptor. Accordingly, A875 cells, which bind proNGF with high affinity¹¹, express high levels of endogenous sortilin and p75^{NTR}

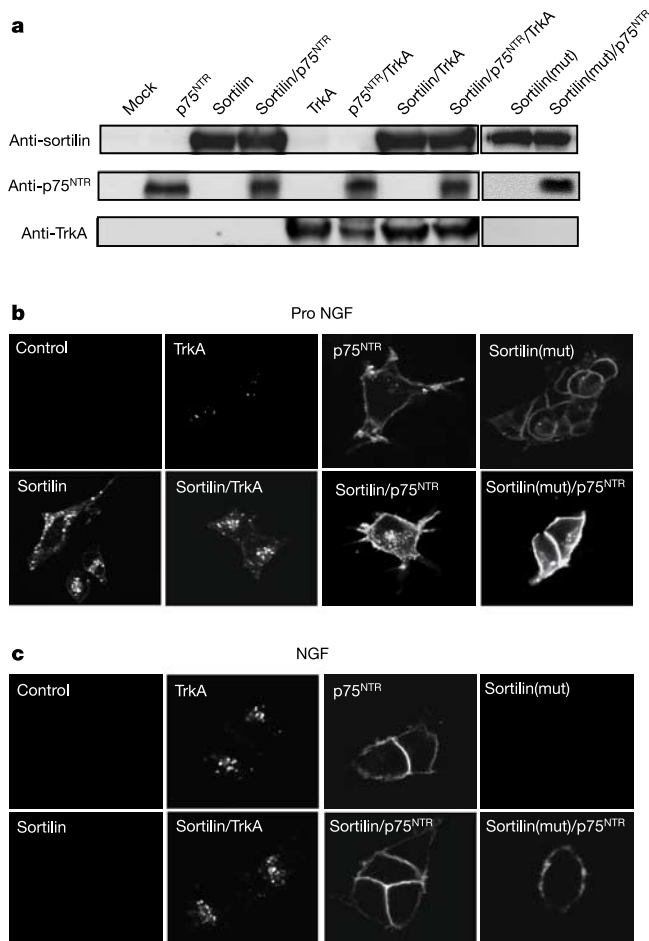


Figure 2 Binding and uptake of proNGF and NGF in 293 cells. **a**, Western blot showing the level of receptor expression in the transfected 293 cells. **b**, **c**, Untransfected cells (control) and cells transfected with the indicated receptors were incubated (37 °C, 45 min) with 50 nM proNGF (**b**) or NGF (**c**) before fixation and staining with anti-NGF antibodies.

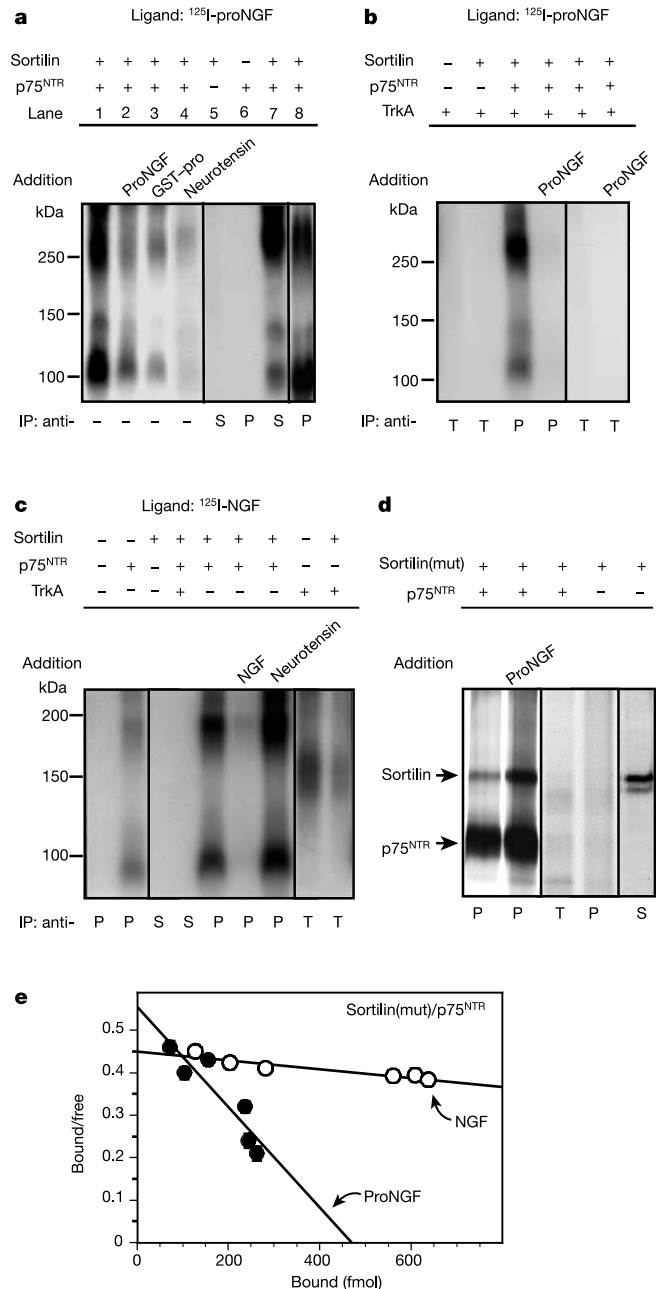


Figure 3 ProNGF-induced formation of heterotrimeric complexes comprising sortilin and p75^{NTR}. **a**–**c**, Crosslinking of ¹²⁵I-labelled proNGF (**a**, **b**) or ¹²⁵I-labelled NGF (**c**) to transfected 293 cells in the presence and absence of excess proNGF, NGF, GST–pro or neurotensin. Crude lysates (–) and immunoprecipitated proteins were subjected to SDS–PAGE and labelled bands were visualized by autoradiography. Antibodies used for immunoprecipitation (IP) of sortilin (S), p75^{NTR} (P) and TrkA (T) are indicated. **d**, Biolabelled cells overexpressing sortilin(mut) and p75^{NTR} were incubated with and without proNGF (25 nM) and treated with a reducible crosslinker. Autoradiographic bands resulting from reducing SDS–PAGE of immunoprecipitates are shown. **e**, Scatchard plot showing binding of radiolabelled NGF (open circles) and proNGF (closed circles) to cells expressing sortilin(mut) and p75^{NTR}.

(Fig. 4a). The results support a model in which proNGF binds to and promotes the formation of a multi-component receptor complex comprising both sortilin and p75^{NTR}.

ProNGF is more efficient than NGF in inducing apoptosis in superior cervical ganglion (SCG) neurons, vascular smooth muscle (SM-11) cells and oligodendrocytes, and in promoting chemotaxis and ligand binding in A875 melanoma cells^{8,11,19}. These cell types all express significant levels of sortilin and p75^{NTR} (Fig. 4a). We found that uptake of proNGF in dissociated SCG cultures was inhibited by the GST-pro protein (data not shown), which selectively inhibits binding to sortilin. Moreover, crosslinking of ¹²⁵I-labelled proNGF to SM-11 cells resulted in labelled adducts of ~110, ~140 and ~240 kDa, similar to those obtained in the p75^{NTR} and sortilin double transfectants (Fig. 4b). These findings suggest that the biological effects of proNGF require coexpression of sortilin and p75^{NTR}. To assess directly whether binding of proNGF to sortilin

regulates biological action, the ability of GST-pro and neurotensin to impair proNGF actions was evaluated. In order to minimize conversion of proNGF into mature NGF, which might introduce a bias by facilitating survival in TrkA-positive cells, a furin-resistant mutant of proNGF¹¹ was used in all subsequent experiments, unless otherwise stated. In SM-11 cells co-expressing p75^{NTR} and sortilin, furin-resistant proNGF was more effective than mature NGF in inducing cell death as assessed by a TdT-mediated dUTP nick end labelling (TUNEL) assay (Fig. 4c). In addition, wild-type proNGF induced apoptosis as effectively as furin-resistant proNGF (data not shown). Co-incubation of proNGF with an excess of neurotensin, or with excess GST-pro, but not GST alone, impaired the induction of cell death and apoptosis by more than 90% (Fig. 4c). Similar findings were obtained in cultured SCG neurons expressing sortilin and p75^{NTR} as well as TrkA. Thus, both GST-pro and neurotensin significantly reduced proNGF-induced apoptosis in SCG neurons,

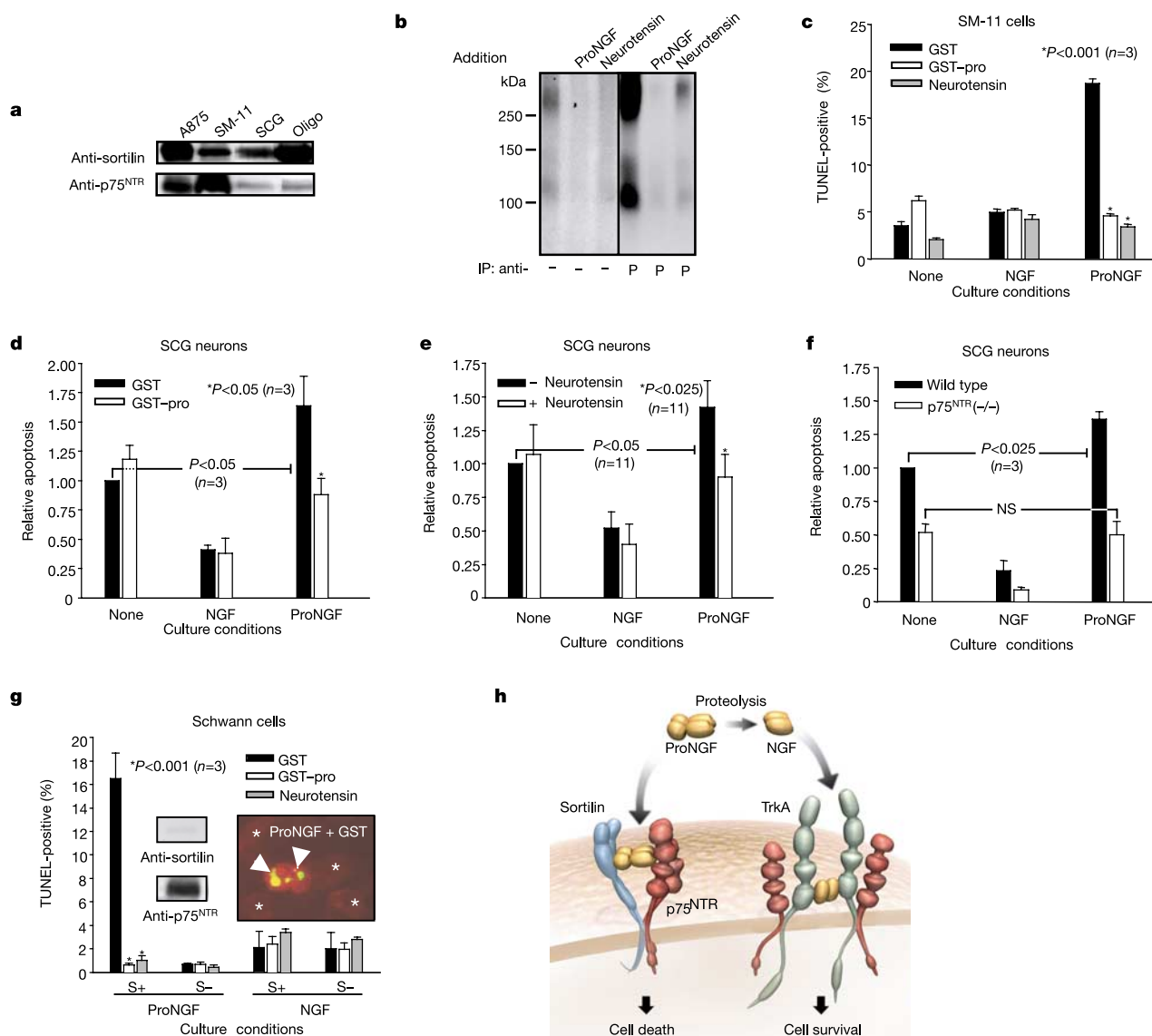


Figure 4 Sortilin is required for the pro-apoptotic action of proNGF. **a**, Receptor expression in proNGF-responsive cell types. **b**, Crosslinking of ¹²⁵I-labelled proNGF to SM-11 cells and inhibition by unlabelled competitors. **c–e**, ProNGF-induced apoptosis (cell type indicated) and its inhibition by GST, GST-pro and neurotensin. The number of apoptotic and total cells counted per condition was 100/~300 (**d** and **f**) and 300/~1,100 (**e**). All values are normalized to apoptosis in the absence of any additions. **f**, Killing of SCG

neurons from wild-type and p75^{NTR} knockout mice. **g**, ProNGF-induced apoptosis of Schwann cells transiently transfected with sortilin. Columns indicate per cent of TUNEL-positive cells among cells that express (S+) or lack (S-) sortilin. Left inset shows western blot of Schwann cell lysate; right inset shows apoptotic nuclei (TUNEL-positive, green) in transfectants expressing sortilin (red staining). Asterisk indicates an untransfected cell. **h**, Schematic model of receptor-complex formation.

whereas neither affected neuronal survival in response to mature NGF or on NGF withdrawal (Fig. 4d, e). As shown in Fig. 4f, p75^{NTR}-deficient neurons from p75^{NTR} knockout mice²⁰ that express sortilin in the absence of p75^{NTR} (data not shown) exhibit NGF-dependent survival, but are resistant to proNGF-induced killing. A similar resistance to proNGF was seen in Schwann cells expressing p75^{NTR} but not sortilin (Fig. 4g); however, after transfection with sortilin, Schwann cells became sensitive to proNGF-induced apoptosis. Approximately 95% of the cells that expressed sortilin (~18% of total) were TUNEL positive, and among all TUNEL-positive cells ~96% expressed both sortilin and p75^{NTR}. This sensitivity to proNGF-induced killing was reversed by co-incubation with GST-pro (Fig. 4g) and neurotensin, which blocked binding of proNGF to sortilin. It follows that both receptors are obligate for the induction of proNGF-mediated cell death, whereas sortilin expression has no impact on NGF responsiveness under these circumstances.

We conclude that proNGF targets and promotes formation of a signalling complex comprising endogenous sortilin and p75^{NTR}, and that both receptors are required for proNGF-mediated apoptosis. In contrast, mature NGF preferentially binds p75^{NTR} and/or TrkA, with sortilin having little or no bearing on NGF-initiated signalling.

Our study indicates that the neurotrophins use not two but three distinct receptor classes to dictate and regulate opposing biological responses of survival and death. We identify sortilin as a biologically important neurotrophin receptor that targets the pro domain of proNGF with high affinity. The present data suggest that sortilin is a required component for transmitting proNGF-dependent death signals via p75^{NTR}. Together with p75^{NTR}, sortilin facilitates the formation of a composite high-affinity binding site for proNGF (Fig. 4h). Thus, sortilin serves as a co-receptor and molecular switch, enabling neurons expressing Trk and p75^{NTR} to respond to a pro-neurotrophin and to initiate pro-apoptotic rather than pro-survival actions. In the absence of sortilin, regulated activity of extracellular proteases may cleave proNGF to mature NGF¹¹, promoting Trk-mediated survival signals (Fig. 4h). In conclusion, NGF-induced neuronal survival and death is far more complicated than previously appreciated, as it depends on an intricate balance between proNGF and mature NGF, as well as on the spatial and temporal expression of three distinct receptors: TrkA, p75^{NTR} and sortilin. As sortilin is but one member of the Vps10p-domain receptor family expressed in the nervous system, future studies should show whether other pro-neurotrophins use related Vps10p-containing receptors to switch biological responsiveness to neurotrophin isoforms. □

Methods

Recombinant proteins and radiolabelling

Human proNGF and mature NGF generated in *Escherichia coli*²¹ were a gift from Scil Proteins GmbH. Radioiodinated ligands^{22,23} (~3,000 d.p.m. fmol⁻¹) were used within 48 h of iodination and their integrity and bioactivity was assayed by SDS-polyacrylamide gel electrophoresis (PAGE), PC12 cell neuritegenesis (NGF) and apoptosis of p75^{NTR}-expressing vascular smooth muscle cells (proNGF). Mature NGF and furin-resistant mouse proNGF were purified from media of transfected 293 cells¹¹. The NGF pro domain (amino acids E19 to R121) was expressed in *E. coli* as a GST fusion protein and purified on glutathione-agarose beads. The luminal domain of sortilin was expressed and purified as described⁵. p75^{NTR}-Fc and TrkA-Fc (fusion proteins) were from R&D Systems.

SPR analysis, equilibrium binding and western blotting

The SPR analysis was performed essentially as described^{5,24}. The receptors were immobilized (at 10–15 µg ml⁻¹) on a CM5 chip and remaining coupling sites were blocked with 1 M ethanolamine. Sample and running buffer was 10 mM HEPES, 150 mM (NH₄)₂SO₄, 1.5 mM CaCl₂, 1 mM EGTA, 0.005% Tween-20 pH 7.4. After each analytic cycle the sensor chip was regenerated in a 10 mM glycine-HCl buffer. The SPR signal was expressed in relative response units (RU); that is, the response obtained in a control flow channel was subtracted. Kinetic parameters were determined using BIAevaluation 3.1 software. Equilibrium-binding studies were performed as described¹². In brief, the cells (2 × 10⁶ ml⁻¹) were incubated (4 °C, 40 min) with radioiodinated NGF or proNGF (2–20 × 10⁻¹⁰ M) in the presence or absence of a 500-fold molar excess of NGF or

neurotensin, respectively. Bound ligand was then separated from free ligand by centrifugation through calf serum. Mean values of triplicates (from two independent experiments) were evaluated using the PRISM program.

Western blotting, after reducing SDS-PAGE, was performed using a rabbit antibody (1:1,000) directed against the pro domain of NGF (amino acids 23–81) of human NGF⁸ and horseradish-peroxidase-conjugated swine anti-rabbit immunoglobulin (Amersham Biosciences).

Transfected cell lines

Parental 293 cells and transfectants expressing p75^{NTR} or TrkA¹² were transfected with wild-type sortilin⁵ or the sortilin(mut) variant impaired in endocytosis (alanine substituted for Y14, L17, L51 and L52 in the cytoplasmic tail)¹⁶, and selected using zeocin. Primary rat Schwann cells²⁵, plated on polylysine-coated plates, were transfected with pcDNA wild-type sortilin⁵ or pcDNA alone using calcium phosphate.

Crosslinking, biolabelling and immunoprecipitation

Cells (2 × 10⁶ ml⁻¹) were incubated (4 °C, 2 h) with radioiodinated proNGF or NGF (400 pM), in the absence or presence of 100 nM unlabelled proNGF, 40 µM neurotensin or 200 nM GST or GST-pro, followed by crosslinking (15 min) with 4 mM 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide and 25 mM DSS (Pierce). Washed cells were subsequently lysed in 1% Nonidet P40 buffer containing protease inhibitors and precipitations were performed using anti-p75^{NTR} and anti-TrkA antisera¹³, and anti-sortilin antibody⁵.

Transfected cells were biolabelled (3–4 h) with L-[³⁵S]cysteine and L-[³⁵S]methionine, then incubated (20 °C, 2 h) with or without 25 nM proNGF and, finally, treated (30 min) with 5 mM of the reducible crosslinker DTSSP (Pierce) before lysis in 1% Triton-X100 buffer containing protease inhibitors^{5,26}. Immunoprecipitation was performed using rabbit anti-p75^{NTR} (number 9993), anti-Trk (Santa Cruz) and anti-sortilin⁵. All precipitated proteins were analysed by reducing SDS-PAGE and were detected by autoradiography.

Induction of apoptosis in various cells

A vascular smooth muscle cell line expressing human p75^{NTR} but not TrkA²⁷ was incubated (16 h) with 2 ng ml⁻¹ of mature NGF or furin-resistant proNGF¹¹ in the presence of 50 nM GST or GST-pro, or 40 µM neurotensin. After fixation, cells were fixed, incubated with 4,6-diamidino-2-phenylindole (DAPI) and subjected to TUNEL analysis (Roche Molecular Biochemicals). Results represent the mean value of three independent experiments performed in triplicate. At least 300 cells per condition were counted.

Dissociated P0–P1 rat SCG neurons²⁸ or mouse SCG neurons obtained from p75^{NTR} knockout mice²⁹ or wild-type littermates were plated on collagen-coated slides and maintained for 5 days in 50 ng ml⁻¹ NGF before use. Replicate cultures were rinsed five times with NGF-free medium and treated with or without the given additives, as indicated. After 36 h SCG cultures were processed for TUNEL analysis and counterstained with anti-neuronal-specific β-tubulin (Tuj1, Covance)¹¹. TUNEL-positive neurons were scored blindly by the observer and at least 100 cells were counted for each culture condition.

Transfected Schwann cells were replated on 8-well slides (NUNC) at 20,000 cells per well. At 48 h after transfection, cells were treated (18 h) with 5 ng ml⁻¹ mature NGF, purified recombinant cleavage-resistant proNGF or diluent alone. After fixation, the cells were stained using mouse anti-sortilin antibody (Transduction Biolabs, anti-NTR3 612101) and rhodamine goat anti-mouse IgG followed by DAPI incubation, and then subjected to TUNEL analysis (Roche Molecular Biochemicals). At least 1,000 cells per condition were counted in a blinded manner, and results are representative of three independent experiments. Where appropriate, statistical significance was determined by Student's *t*-test.

Received 3 November; accepted 23 December 2003; doi:10.1038/nature02319.

- Petersen, C. M. *et al.* Molecular identification of a novel candidate sorting receptor purified from human brain by receptor-associated protein affinity chromatography. *J. Biol. Chem.* **272**, 3599–3605 (1997).
- Hermey, G. *et al.* Characterization of sorCS1, an alternatively spliced receptor with completely different cytoplasmic domains that mediate different trafficking in cells. *J. Biol. Chem.* **278**, 7390–7396 (2003).
- Jacobsen, L. *et al.* Activation and functional characterization of the mosaic receptor SorLA/LR11. *J. Biol. Chem.* **276**, 22788–22796 (2001).
- Mazella, J. *et al.* The 100-kDa neurotensin receptor is gp95/sortilin, a non-G-protein-coupled receptor. *J. Biol. Chem.* **273**, 26273–26276 (1998).
- Munck, P. C. *et al.* Propeptide cleavage conditions sortilin/neurotensin receptor-3 for ligand binding. *EMBO J.* **18**, 595–604 (1999).
- Sarret, P. *et al.* Distribution of NTS3 receptor/sortilin mRNA and protein in the rat central nervous system. *J. Comp. Neurol.* **461**, 483–505 (2003).
- Hermans-Borgmeyer, I., Hermey, G., Nykjaer, A. & Schaller, C. Expression of the 100-kDa neurotensin receptor sortilin during mouse embryonal development. *Brain Res. Mol. Brain Res.* **65**, 216–219 (1999).
- Beattie, M. S. *et al.* ProNGF induces p75-mediated death of oligodendrocytes following spinal cord injury. *Neuron* **36**, 375–386 (2002).
- Hasan, W., Pedchenko, T., Krizsan-Agbas, D., Baum, L. & Smith, P. G. Sympathetic neurons synthesize and secrete pro-nerve growth factor protein. *J. Neurobiol.* **57**, 38–53 (2003).
- Chao, M. V. Neurotrophins and their receptors: a convergence point for many signalling pathways. *Nature Rev. Neurosci.* **4**, 299–309 (2003).
- Lee, R., Keramian, P., Teng, K. K. & Hempstead, B. L. Regulation of cell survival by secreted proneurotrophins. *Science* **294**, 1945–1948 (2001).
- Esposito, D. *et al.* The cytoplasmic and transmembrane domains of the p75 and Trk A receptors regulate high affinity binding to nerve growth factor. *J. Biol. Chem.* **276**, 32687–32695 (2001).

13. Mahadeo, D., Kaplan, L., Chao, M. V. & Hempstead, B. L. High affinity nerve growth factor binding displays a faster rate of association than p140trk binding. Implications for multi-subunit polypeptide receptors. *J. Biol. Chem.* **269**, 6884–6891 (1994).
14. Fahnstock, M., Michalski, B., Xu, B. & Coughlin, M. D. The precursor pro-nerve growth factor is the predominant form of nerve growth factor in brain and is increased in Alzheimer's disease. *Mol. Cell. Neurosci.* **18**, 210–220 (2001).
15. Heymach, J. V. Jr & Shooter, E. M. The biosynthesis of neurotrophin heterodimers by transfected mammalian cells. *J. Biol. Chem.* **270**, 12297–12304 (1995).
16. Nielsen, M. S. *et al.* The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20**, 2180–2190 (2001).
17. Gargano, N., Levi, A. & Alema, S. Modulation of nerve growth factor internalization by direct interaction between p75 and TrkA receptors. *J. Neurosci. Res.* **50**, 1–12 (1997).
18. Bronfman, F. C., Tcherpakov, M., Jovin, T. M. & Fainzilber, M. Ligand-induced internalization of the p75 neurotrophin receptor: a slow route to the signaling endosome. *J. Neurosci.* **23**, 3209–3220 (2003).
19. Shonukan, O., Bagayogo, I., McCrea, P., Chao, M. & Hempstead, B. Neurotrophin-induced melanoma cell migration is mediated through the actin-bundling protein fascin. *Oncogene* **22**, 3616–3623 (2003).
20. Lee, K. F. *et al.* Targeted mutation of the gene encoding the low affinity NGF receptor p75 leads to deficits in the peripheral sensory nervous system. *Cell* **69**, 737–749 (1992).
21. Rattenholl, A. *et al.* The pro-sequence facilitates folding of human nerve growth factor from *Escherichia coli* inclusion bodies. *Eur. J. Biochem.* **268**, 3296–3303 (2001).
22. Hempstead, B. L., Schleifer, L. S. & Chao, M. V. Expression of functional nerve growth factor receptors after gene transfer. *Science* **243**, 373–375 (1989).
23. Hempstead, B. L., Martin-Zanca, D., Kaplan, D. R., Parada, L. F. & Chao, M. V. High-affinity NGF binding requires coexpression of the trk proto-oncogene and the low-affinity NGF receptor. *Nature* **350**, 678–683 (1991).
24. Nykjaer, A. *et al.* Cubilin dysfunction causes abnormal metabolism of the steroid hormone 25(OH) vitamin D(3). *Proc. Natl Acad. Sci. USA* **98**, 13895–13900 (2001).
25. Einheber, S., Milner, T. A., Giancotti, F. & Salzer, J. L. Axonal regulation of Schwann cell integrin expression suggests a role for $\alpha 6 \beta 4$ in myelination. *J. Cell Biol.* **123**, 1223–1236 (1993).
26. Nykjaer, A. *et al.* Mannose 6-phosphate/insulin-like growth factor-II receptor targets the urokinase receptor to lysosomes via a novel binding interaction. *J. Cell Biol.* **141**, 815–828 (1998).
27. Wang, S. *et al.* p75(NTR) mediates neurotrophin-induced apoptosis of vascular smooth muscle cells. *Am. J. Pathol.* **157**, 1247–1258 (2000).
28. Mitsui, C., Sakai, K., Ninomiya, T. & Koike, T. Involvement of TLCK-sensitive serine protease in colchicine-induced cell death of sympathetic neurons in culture. *J. Neurosci. Res.* **66**, 601–611 (2001).
29. Bamji, S. X. *et al.* The p75 neurotrophin receptor mediates neuronal apoptosis and is essential for naturally occurring sympathetic neuron death. *J. Cell Biol.* **140**, 911–923 (1998).

Acknowledgements We thank M. V. Chao and G. R. Lewin for valuable discussions. J. Salzer, R. Kraemer and P. Fischer are acknowledged for reagents and advice, and S. Tevar for assistance in p75^{NTR} mice genotyping. This work was supported by the Novo Nordisk Foundation, The Danish Medical Research Council, The Carlsberg Foundation (A.N. and C.M.P.) and the NIH (B.L.H. and R.L.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.N. (an@biokemi.au.dk).

The cytoplasmic body component TRIM5 α restricts HIV-1 infection in Old World monkeys

Matthew Stremlau¹, Christopher M. Owens¹, Michel J. Perron¹, Michael Kiessling¹, Patrick Autissier² & Joseph Sodroski^{1,3}

¹Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Department of Pathology, Division of AIDS, and ²Division of Viral Pathogenesis, Beth Israel Deaconess Medical Center, Department of Medicine, Division of AIDS, Harvard Medical School, Boston, Massachusetts 02115, USA

³Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA

Host cell barriers to the early phase of immunodeficiency virus replication explain the current distribution of these viruses among human and non-human primate species^{1–4}. Human immunodeficiency virus type 1 (HIV-1), the cause of acquired immunodeficiency syndrome (AIDS) in humans, efficiently enters the cells of Old World monkeys but encounters a block before reverse transcription^{2–4}. This species-specific restriction acts on the incoming HIV-1 capsid^{5–7} and is mediated by a

dominant repressive factor^{7–9}. Here we identify TRIM5 α , a component of cytoplasmic bodies, as the blocking factor. HIV-1 infection is restricted more efficiently by rhesus monkey TRIM5 α than by human TRIM5 α . The simian immunodeficiency virus, which naturally infects Old World monkeys¹⁰, is less susceptible to the TRIM5 α -mediated block than is HIV-1, and this difference in susceptibility is due to the viral capsid. The early block to HIV-1 infection in monkey cells is relieved by interference with TRIM5 α expression. Our studies identify TRIM5 α as a species-specific mediator of innate cellular resistance to HIV-1 and reveal host cell components that modulate the uncoating of a retroviral capsid.

Recombinant HIV-1 expressing green fluorescent protein and pseudotyped with the vesicular stomatitis virus (VSV) G glycoprotein (denoted HIV-1–GFP) can efficiently infect the cells of many mammalian species including humans, but not those of Old World monkeys^{4–9}. Here we used a murine leukaemia virus vector to transduce human HeLa cells, which are susceptible to HIV-1–GFP infection, with a complementary DNA library prepared from primary rhesus monkey lung fibroblasts (PRL cells). Two independent HeLa clones resistant to HIV-1–GFP infection, but susceptible to infection with recombinant simian immunodeficiency virus (SIV–GFP) or murine leukaemia virus (MLV–GFP), were identified in a screen (Methods).

The only monkey cDNA insert common to both HIV-1–GFP-resistant clones was predicted to encode TRIM5 α , a member of the tripartite motif (TRIM) family of proteins containing RING domains, B-boxes and coiled coils^{11–13}. TRIM5 α also contains a carboxy-terminal B30.2 (SPRY) domain not found in the other TRIM5 isoforms (ref. 13 and Fig. 1a). The natural functions of TRIM5 α , or of the cytoplasmic bodies in which the TRIM5 proteins localize^{13,14}, are unknown. One TRIM5 isoform has been shown to have ubiquitin ligase activity typical of RING-containing proteins¹⁴. TRIM5 proteins are expressed constitutively in many tissues¹³, consistent with the pattern of expression expected for the HIV-1-blocking factor in monkeys⁴.

HeLa cells stably expressing rhesus monkey TRIM5 α (TRIM5 α_{rh}) and control HeLa cells containing empty vector were incubated with different amounts of recombinant HIV-1–GFP, SIV–GFP and MLV–GFP. Expression of TRIM5 α_{rh} resulted in a marked inhibition of infection by HIV-1–GFP, whereas MLV–GFP infected control and TRIM5 α_{rh} -expressing HeLa cells equivalently (Fig. 1b, c). TRIM5 α_{rh} inhibited infection by SIV–GFP less efficiently than that by HIV-1–GFP (Fig. 1c). Stable TRIM5 α_{rh} expression also inhibited the replication of infectious HIV-1 in HeLa-CD4 cells, which express the receptors for HIV-1 (ref. 15 and Fig. 1d). The replication of a simian–human immunodeficiency virus (SHIV) chimera, which contains core proteins (including the capsid protein) of SIV_{mac} (ref. 16), was not inhibited in these TRIM5 α_{rh} -expressing cells. When the infections were done with eightfold more HIV-1 and SHIV, similar results were obtained (Supplementary Information). We conclude that expression of TRIM5 α_{rh} specifically and efficiently blocks infection by HIV-1, and exerts a slight inhibitory effect on infection by SIV_{mac}.

To investigate the viral target of the TRIM5 α_{rh} -mediated restriction, HeLa cells expressing TRIM5 α_{rh} or control HeLa cells were incubated with recombinant HIV-1–GFP, SIV–GFP, SIV(HCA-p2)–GFP or HIV(SCA)–GFP. SIV(HCA-p2)–GFP is identical to SIV–GFP, except that the SIV capsid and adjacent p2 sequences have been replaced by those of HIV-1 (ref. 17), and SIV(HCA-p2)–GFP has been shown to be susceptible to the block in Old World monkey cells^{5,7}. HIV(SCA)–GFP is identical to HIV-1–GFP, except that most of the capsid protein has been replaced by that of SIV⁵, and HIV(SCA)–GFP has been shown to be less susceptible than HIV-1 to the block in Old World monkey cells⁵. We found that HIV-1–GFP and SIV(HCA-p2)–GFP infections were restricted to the same extent in TRIM5 α_{rh} -expressing HeLa cells, whereas infec-

tions by SIV–GFP and HIV(SCA)–GFP were less restricted in these cells (Fig. 1e). We conclude that capsid sequences influence viral susceptibility to the TRIM5 α_{rh} -mediated restriction.

To determine the level at which HIV-1 infection is blocked by TRIM5 α_{rh} , we used a real-time polymerase chain reaction (PCR) assay to detect viral cDNA at various times after incubating VSV G-pseudotyped HIV-1 with either HeLa cells expressing TRIM5 α_{rh}

or control HeLa cells (Fig. 2). After 2 h, the levels of early reverse transcripts were comparable in TRIM5 α_{rh} -expressing and control HeLa cells. At later time points, both early and late reverse transcripts were barely detectable in TRIM5 α_{rh} -expressing cells; by contrast, both early and late viral cDNAs were abundant in control cells. As the VSV G glycoproteins support entry into these cells equivalently (see MLV–GFP infection in Fig. 1b, c), these data

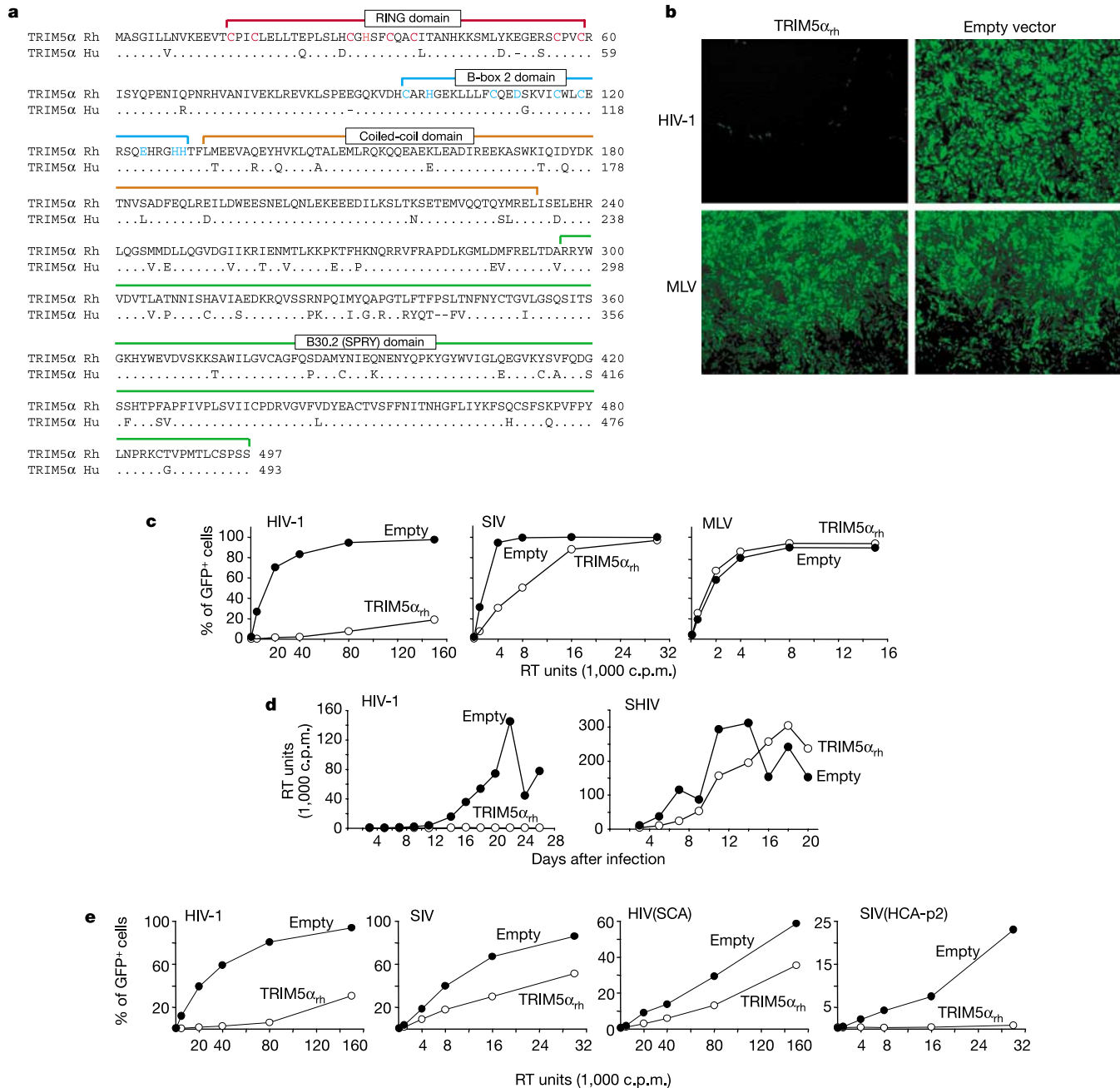


Figure 1 Rhesus monkey TRIM5 α_{rh} preferentially blocks HIV-1 infection. **a**, Alignment of amino acid sequences of rhesus monkey TRIM5 α_{rh} and human TRIM5 α_{hu} (predicted from the sequences of the cDNAs used in this study), with key domains indicated. Residues that are conserved among RING and B-box 2 domains are coloured. **b**, HeLa cells transduced with pLPCX–TRIM5 $\alpha_{rh}(fl)$ vector expressing TRIM5 α_{rh} or an empty control vector (pLPCX) were selected in medium containing puromycin and incubated with 2×10^4 RT units of HIV-1–GFP or 8×10^3 RT units of MLV–GFP. Infected, GFP-positive cells were visualized by fluorescence microscopy. **c**, HeLa cells transduced with pLPCX–TRIM5 $\alpha_{rh}(fl)$ expressing TRIM5 α_{rh} (open circles) or empty vector (filled circles) were incubated with

various amounts of the indicated viruses expressing GFP. Infected, GFP-positive cells were counted by FACS. **d**, HeLa-CD4 cells transduced with pLPCX–TRIM5 $\alpha_{rh}(cds)$ vector expressing TRIM5 α_{rh} (open circles) or empty vector (filled circles) were incubated with 1×10^4 RT units of HIV-1 or SHIV as indicated. **e**, HeLa cells transduced with pLPCX–TRIM5 $\alpha_{rh}(fl)$ vector expressing TRIM5 α_{rh} (open circles) or empty vector (filled circles) were exposed to the indicated GFP-expressing viruses. Infected, GFP-positive cells were counted by FACS. Results in **c** and **e** are typical of those obtained in at least three independent experiments.

indicate that early events after HIV-1 entry are impaired in cells expressing TRIM5 α_{rh} . Directly or indirectly, TRIM5 α_{rh} disrupts viral cDNA synthesis or accelerates its decay.

The predicted sequences (Fig. 1a) of TRIM5 α_{rh} and its human orthologue, TRIM5 α_{hu} , show differences that might account for the species-specific nature of the early block to HIV-1 infection. To test this hypothesis, TRIM5 α_{rh} -HA and TRIM5 α_{hu} -HA proteins, containing C-terminal epitope tags from influenza haemagglutinin,

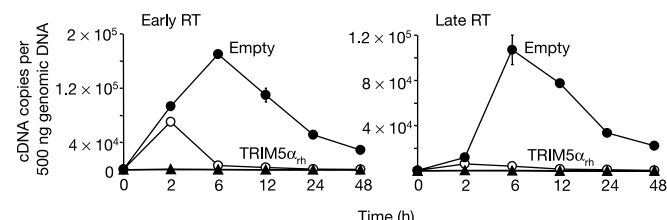


Figure 2 TRIM5 α_{rh} blocks HIV-1 infection before or during early reverse transcription. HeLa cells stably transduced with pLPCX-TRIM5 α_{rh} (fl) vector expressing TRIM5 α_{rh} (open circles) or with empty pLPCX vector (filled circles) were incubated with VSV G-pseudotyped, DNase-treated HIV-1-GFP. HeLa cells transduced with empty vector were also incubated with HIV-1-GFP viruses lacking envelope glycoproteins (filled triangles). Target cell DNA was isolated at the indicated times and used to detect early (left) and late (right) reverse transcripts. Data are the means $\pm$ s.d. from duplicate experiments.

were expressed stably in HeLa cells. Similar levels of TRIM5 α_{rh} -HA and TRIM5 α_{hu} -HA were expressed in HeLa cells (Fig. 3a). Neither TRIM5 α_{rh} -HA nor TRIM5 α_{hu} -HA affected the efficiency with which HeLa cells were infected by MLV-GFP (Fig. 3b). TRIM5 α_{hu} -HA inhibited HIV-1-GFP infection less efficiently than did TRIM5 α_{rh} -HA. Parallel experiments with TRIM5 α proteins lacking the epitope tags verified that fusion of HA to the C terminus of TRIM5 α_{rh} and TRIM5 α_{hu} did not affect the efficiency with which these proteins suppressed HIV-1 infection (Supplementary Information). By contrast, the modest inhibitory effect of TRIM5 α_{rh} on SIV-GFP infection was not seen for the epitope-tagged TRIM5 α_{rh} -HA protein (Fig. 3b). Neither TRIM5 α_{hu} nor TRIM5 α_{hu} -HA exerted significant effects on SIV-GFP infection. Thus, differences in the TRIM5 α proteins of humans and monkeys probably contribute to the species-specific differences in susceptibility to HIV-1 infection.

We examined the ability of rhesus monkey TRIM5 variants to inhibit HIV-1 infection. Differential splicing of *TRIM5* transcripts results in the production of several isoforms that lack the TRIM5 α C terminus¹³. For example, the first 300 amino acid residues of the TRIM5 γ_{rh} isoform are identical to those of TRIM5 α_{rh} , but the TRIM5 γ_{rh} sequence diverges and stops thereafter. To examine the contribution of the C-terminal B30.2 (SPRY) domain of TRIM5 α_{rh} to the inhibition of HIV-1 infection, we tested an epitope-tagged version of the TRIM5 γ_{rh} isoform. We also examined whether an intact RING domain is required for TRIM5 α_{rh} -mediated inhibition of HIV-1 infection. Studies of proteins containing RING domains, including TRIM5 δ , indicate that alteration of

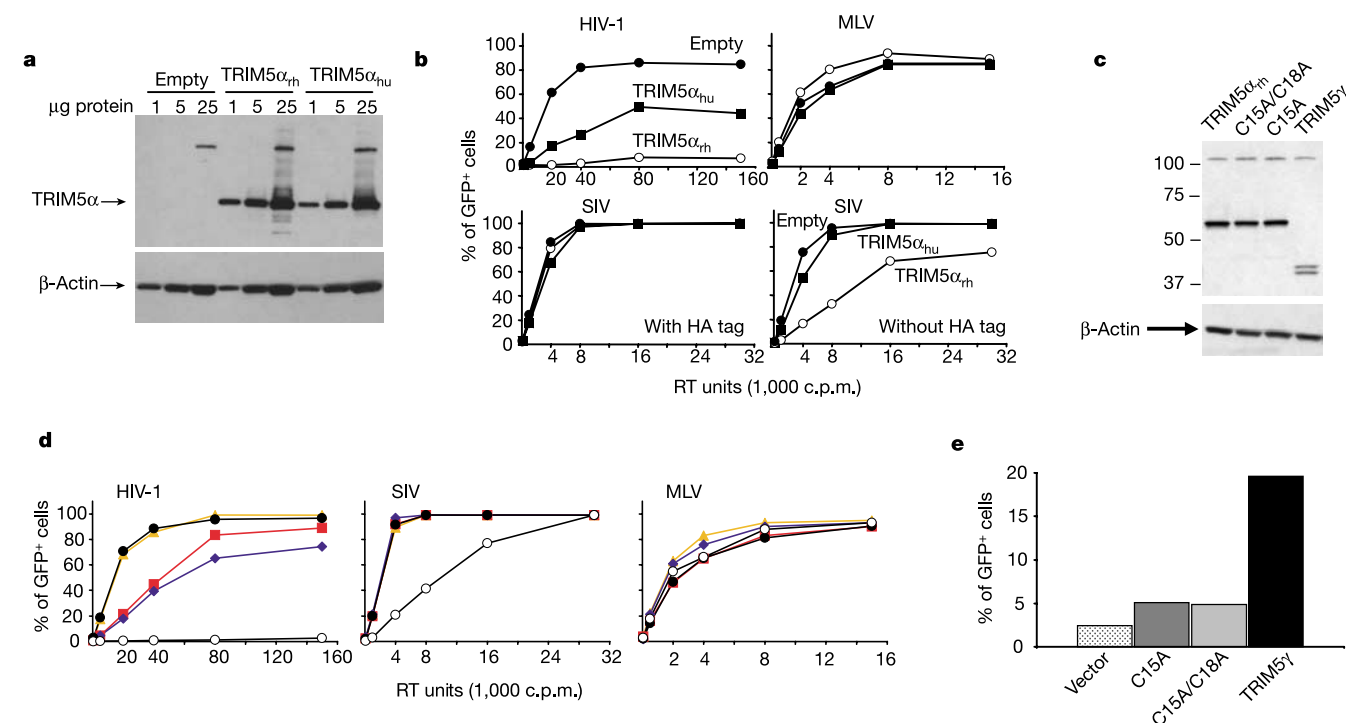


Figure 3 HIV-1 infection is blocked less efficiently by TRIM5 variants than by TRIM5 α_{rh} . **a**, The indicated amounts of total protein from lysates of HeLa cells stably expressing human TRIM5 α_{hu} -HA or rhesus monkey TRIM5 α_{rh} -HA, or transduced with empty pLPCX vector, were subjected to western blotting with an antibody against HA. **b**, HeLa cells stably expressing TRIM5 α_{hu} -HA (filled squares) or TRIM5 α_{rh} -HA (open circles), or transduced with empty vector (filled circles), were incubated with the indicated amounts of GFP-expressing viruses. Infected, GFP-positive cells were counted by FACS. One set of SIV-GFP infections was carried out on cells expressing TRIM5 α_{hu} or TRIM5 α_{rh} proteins without HA epitope tags. **c**, Lysates (normalized for protein) from HeLa cells transfected with 4 μ g of pLPCX-TRIM5 α_{rh} (cds) variants expressing HA-tagged versions of TRIM5 α_{rh} ,

TRIM5 α_{rh} -C15A, TRIM5 α_{rh} -C15A/C18A or TRIM5 γ were subjected to western blotting with an antibody against HA. **d**, HeLa cells stably expressing untagged versions of TRIM5 α_{rh} (open circles), TRIM5 α_{rh} -C15A (red squares), TRIM5 α_{rh} -C15A/C18A (blue diamonds) or TRIM5 γ (orange triangles), or transduced with empty vector (filled circles), were incubated with the indicated GFP-expressing viruses. Infected, GFP-positive cells were counted by FACS. **e**, PRL cells were transduced with empty vector, or vectors expressing TRIM5 α_{rh} -C15A, TRIM5 α_{rh} -C15A/C18A or TRIM5 γ , and infected with 4×10^4 RT units of HIV-1-GFP. Infected, GFP-positive cells were counted by FACS. Results in **d** and **e** are typical of those obtained in at least three independent experiments.

one or both of the two most amino-terminal cysteines inactivates ubiquitin ligase activity^{14,18}. Thus, we created and tested two HA-tagged TRIM5 α_{rh} mutants, TRIM5 α_{rh} -C15A and TRIM5 α_{rh} -C15A/C18A. Expression of TRIM5 γ_{rh} and the RING domain cysteine mutants was comparable to that of wild-type TRIM5 α_{rh} (Fig. 3c). TRIM5 γ_{rh} did not inhibit HIV-1-GFP infection and, compared with wild-type TRIM5 α_{rh} , TRIM5 α_{rh} -C15A and TRIM5 α_{rh} -C15A/C18A were significantly less effective in inhibiting HIV-1-GFP infection (Fig. 3d). We conclude that both the RING domain at the N terminus and the B30.2 (SPRY) domain at the C terminus of TRIM5 α_{rh} contribute to the HIV-1-inhibitory activity of this protein.

The TRIM5 γ_{rh} , TRIM5 α_{rh} -C15A and TRIM5 α_{rh} -C15A/C18A variants were also examined for dominant-negative activity in

relieving the block to HIV-1-GFP infection in PRL cells. We found that HIV-1-GFP infected PRL cells expressing TRIM5 γ_{rh} more efficiently than it infected control PRL cells transduced with empty vector or PRL cells expressing the RING domain cysteine mutants of TRIM5 α_{rh} (Fig. 3e). Thus, TRIM5 γ_{rh} acts in a dominant-negative manner to suppress the restriction to HIV-1 in PRL cells.

To determine whether TRIM5 α_{rh} is necessary for the restriction to HIV-1 infection in Old World monkey cells, short interfering RNAs (siRNAs) directed against TRIM5 α_{rh} were used to down-regulate its expression. Several different siRNAs directed against TRIM5 α_{rh} , including siRNA 1 and siRNA 3 (Methods), reduced TRIM5 α_{rh} -HA expression in HeLa cells (Fig. 4a and data not shown). As expected, siRNA 3 did not inhibit the expression of TRIM5 α_{rh} -escape, a wild-type HA-tagged protein encoded by an

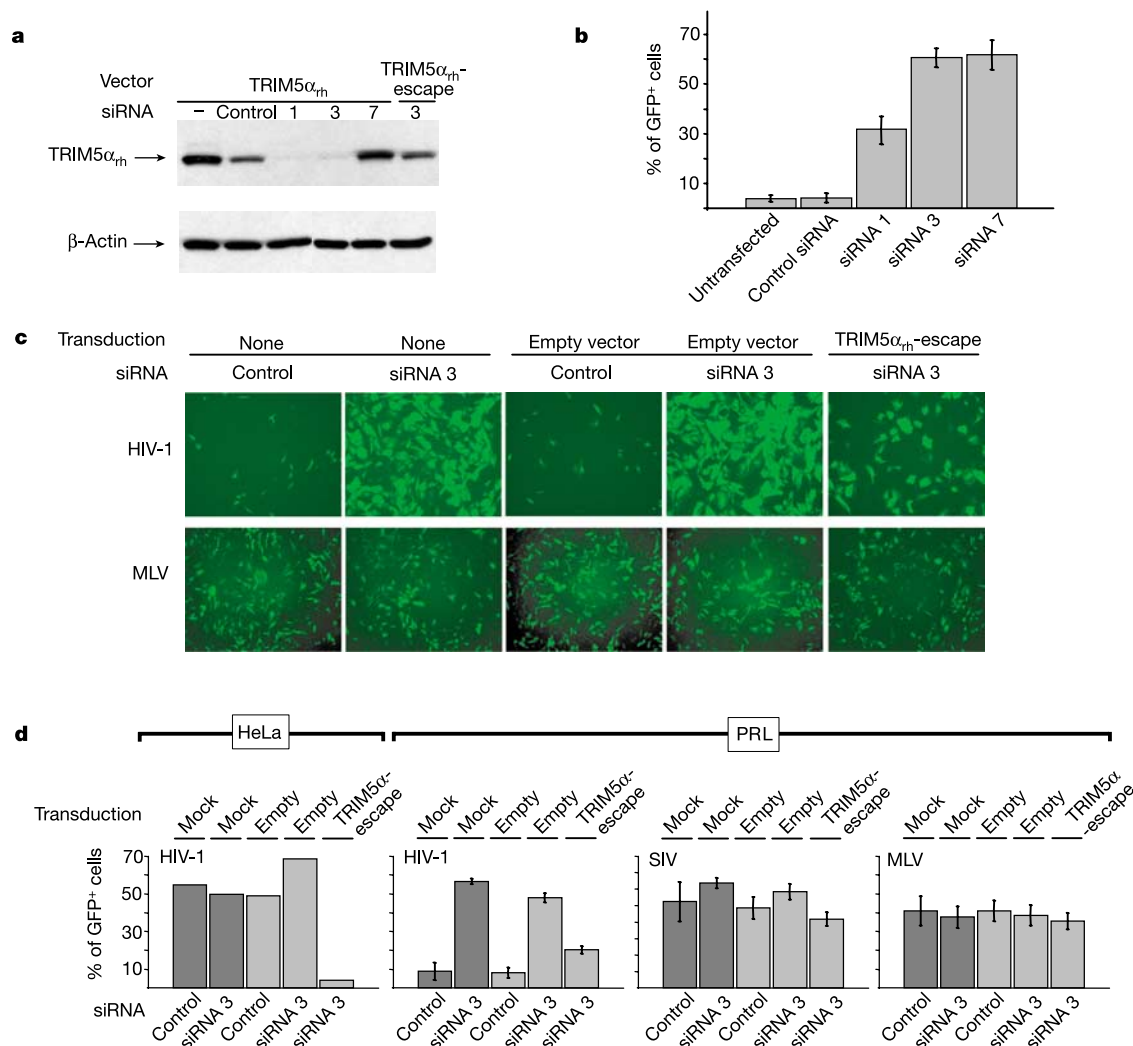


Figure 4 TRIM5 α_{rh} is essential for the block to HIV-1 infection in PRL cells. **a**, HeLa cells were co-transfected with 4 μ g of pLPCX-TRIM5 α_{rh} (cds) variants expressing HA-tagged TRIM5 α_{rh} or TRIM5 α_{rh} -escape and 120 nM of the indicated siRNAs. Cell lysates were normalized for protein and subjected to western blotting with antibodies against HA or β -actin. Note that the siRNA 7 target sequence, which is located in the 3'-UTR of the natural TRIM5 α_{rh} mRNA, is missing in the pLPCX-TRIM5 α_{rh} (cds) plasmid used to express the HA-tagged TRIM5 α_{rh} ; siRNA 7 therefore serves as a negative control. **b**, The indicated siRNAs were transfected into PRL cells, which were then incubated with HIV-1-GFP. Infected GFP-positive cells were counted by FACS. Data are the means $\pm$ s.d. from three independent experiments. Note that the wild-type TRIM5 α_{rh} mRNA in PRL cells contains the target sequences for siRNAs 1, 3 and 7; therefore, all three siRNAs are expected to

interfere with TRIM5 α_{rh} expression. **c**, Untreated PRL cells and PRL cells transduced with empty vector (pLPCX) or a vector expressing TRIM5 α_{rh} -escape were transfected with 120 nM of either a control siRNA or siRNA 3. The cells were incubated with HIV-1-GFP or MLV-GFP and after 48 h were visualized by fluorescence microscopy. Original magnification, $\times 4$ for all images. **d**, The indicated cells, either untreated or transduced with empty vector or a vector expressing TRIM5 α_{rh} -escape, were transfected with 120 nM of control siRNA or siRNA 3 and then incubated with the indicated GFP-expressing viruses. Infected, GFP-positive cells were counted by FACS. Data for the PRL cells are the means $\pm$ s.d. from three independent experiments; the experiment in HeLa cells was done twice with similar results.

expression vector with six silent mutations in the siRNA 3 target sequence (Fig. 4a). Transfection of PRL cells with the TRIM5 α_{rh} -directed siRNAs, but not with a control siRNA, resulted in a marked increase in the efficiency of HIV-1-GFP infection (Fig. 4b–d). The siRNAs targeting TRIM5 α_{rh} did not affect MLV-GFP infection (Fig. 4c, d, and data not shown). The increase observed after transfection of siRNA 3 was diminished when the PRL cells were transduced with a vector expressing TRIM5 α_{rh} -escape (Fig. 4c, d). Although the restriction to HIV-1 infection is not as strong in LLC-MK2 cells as it is in PRL cells, siRNA directed against TRIM5 α_{rh} also increased the efficiency of HIV-1-GFP infection in LLC-MK2 cells (Supplementary Information). We conclude that TRIM5 α_{rh} is an essential factor for the early block to HIV-1 in Old World monkey cells.

The maintenance of a strong block to HIV-1 in Old World monkeys implies a selective advantage, presumably imposed by the presence of HIV-1-like viruses during the evolution of this primate lineage. Here, TRIM5 α_{rh} has been identified as a key mediator of the monkey cell restriction to HIV-1. Expression of TRIM5 α_{rh} was necessary for the early block to HIV-1 in monkey cells and sufficient for establishing such a block in human cells. Differences in the expression level, or polymorphisms in TRIM5 α_{rh} or its associated cofactors, may account for the different degrees of HIV-1 restriction observed in several established Old World monkey cell lines⁴.

The mechanism of the TRIM5 α_{rh} -mediated block to HIV-1 infection requires further investigation, but the importance of the TRIM5 α_{rh} RING domain and the capsid specificity of the restriction indicate that TRIM5 α_{rh} may directly bind and ubiquitinate the HIV-1 capsid. HIV-1 capsid mutants¹⁹ that bind the monkey cell restriction factor efficiently, but still infect monkey cells, might be resistant to the downstream effects of TRIM5 α binding, including ubiquitination. Capsid modification by ubiquitin could adversely affect uncoating, which occurs soon after HIV-1 entry²⁰. Studies of HIV-1 Gag mutants suggest that capsid disassembly shows precise requirements; that is, both increases and decreases in capsid stability are detrimental to HIV-1 replication²¹.

Although human and Old World monkey cells are susceptible to infection by HIV-1 and SIV_{mac}, respectively, the TRIM5 α proteins from these cells showed some ability to repress the infecting virus. Variations in expression or polymorphisms in TRIM5 α may thus influence the course of natural infection by these viruses. The treatment of human target cells with proteasome inhibitors results in an increase in the early phase of HIV-1 infection²², which suggests that HIV-1-suppressive processes involving ubiquitination may be operative in human cells. TRIM5 α_{hu} was less effective in suppressing HIV-1 and SIV_{mac} infection than was TRIM5 α_{rh} . Notably, it has been suggested that HIV-1 capsids bind the Old World monkey restriction factor more efficiently than do SIV_{mac} capsids^{7–9}. Apparently, each virus has evolved in its natural host to achieve an acceptably low level of TRIM5 α interaction. Vigorous, detrimental capsid disassembly may result when efficiently binding capsids, like those of HIV-1, encounter more effective TRIM5 α proteins, like those expressed in simian cells.

The TRIM proteins constitute a large family whose members exhibit diverse functions and cellular locations¹³. A nuclear TRIM protein, PML, has been reported to inhibit the replication or expression of various viruses, including HIV-1 (refs 23–26). The function of cytoplasmic bodies, which contain a subset of TRIM family members¹³, is unknown; our results suggest that cytoplasmic bodies may contribute to innate cellular resistance to viruses. The induction of expression of some TRIM proteins^{26–28} by interferon and the expansion of this gene family in parallel with the metazoan lineage are consistent with this idea.

Understanding the early species-specific restrictions to HIV-1 replication, in conjunction with the late block owing to APO-BEC3G²⁹, may suggest approaches to the development of animal models of HIV-1 infection. Furthermore, insight into the uncoating

process, hitherto a poorly understood aspect of the retroviral life cycle, should facilitate intervention. □

Methods

Screen for HIV-1-resistant cells

A PRL⁴ cDNA library (3.2×10^6 independent clones) was inserted into the pLIB vector (Clontech) and used to transduce 3×10^6 HeLa cells. After 3 d, 6×10^6 transduced cells were reseeded in batches of 5×10^5 cells in 10-cm dishes and incubated with sufficient HIV-1-GFP to infect at least 99% of the cells. About 0.5% of the cells were selected for absence of fluorescence by a FACS Vantage SE cell sorter (Becton Dickinson). Collected cells were allowed to grow into 500-cell colonies and subjected to a second round of HIV-1-GFP infection at a high multiplicity of infection. GFP-negative colonies were identified by fluorescence microscopy, cloned and expanded.

We selected 313 HeLa clones from seven sequential screens and tested them for susceptibility to HIV-1-GFP and SIV-GFP. Two clones with a selective block to HIV-1-GFP were identified. Eleven cDNA inserts from these two clones were recovered by PCR amplification of genomic DNA samples with oligonucleotide primers specific for the pLIB vector and the following conditions: 1 cycle at 95 °C for 1 min, 40 cycles of 95 °C for 1 min, 68 °C for 1 min and 72 °C for 5 min, and then 1 cycle at 72 °C for 10 min. Each cDNA was subcloned into the *EcoRI* and *Clal* restriction sites of the pLPCX vector (Clontech) and sequenced. The cDNA encoding TRIM5 α_{rh} was the only monkey cDNA present in both HIV-1-resistant HeLa clones and was the only cDNA subsequently confirmed to inhibit HIV-1 infection.

Creation of cells stably expressing TRIM5 α variants

We co-transfected pLPCX vectors containing TRIM5 cDNAs or control empty pLPCX vectors into 293T cells by using pVPack-GP and pVPack-VSV-G (both from Stratagene) packaging plasmids. The resulting virus particles were used to transduce 5×10^5 HeLa cells in the presence of $5 \mu\text{g ml}^{-1}$ polybrene, and the cells were subjected to selection in $1 \mu\text{g ml}^{-1}$ puromycin (Sigma). Cells were transduced with either a pLPCX-TRIM5 α_{rh} (fl) vector containing full-length TRIM5 α_{rh} cDNA, which includes the 5' and 3'-untranslated regions (UTRs), or a pLPCX-TRIM5 α_{rh} (cds) vector containing only the amino acid coding sequence of the TRIM5 α_{rh} cDNA.

Infection with viruses expressing GFP

We prepared HIV-1-GFP, SIV-GFP, HIV(SCA)-GFP and SIV(HCA-p2)-GFP viruses as described^{4,5,17}, and MLV-GFP by co-transfecting 293T cells with $15 \mu\text{g}$ of pFB-hrGFP, $15 \mu\text{g}$ of pVPack-GP and $4 \mu\text{g}$ of pVPack-VSV-G (all from Stratagene). HIV and SIV viral stocks were quantified by measuring reverse transcriptase (RT) activity as described¹⁶. MLV RT activity was determined by the same procedure except that 20 mM MnCl₂ was used instead of MgCl₂. For infections, 3×10^4 HeLa cells or 2×10^4 PRL cells seeded in 24-well plates were incubated in the presence of virus for 24 h. Cells were washed and returned to culture for 48 h, and then subjected to FACS analysis with a FACScan (Becton Dickinson).

Infection with replication-competent virus

Stocks of replication-competent HIV-1_{HXBc2} and SHIV_{HXBc2} (ref. 16) were prepared from supernatants of 293T cells transfected with the respective proviral clones. Spreading infections were initiated with stocks normalized according to RT activity, and replication was monitored over time by analysing culture supernatants for RT activity.

Quantitative real-time PCR

Virus stocks derived from the transfection of 293T cells were treated with 50 U ml^{-1} Turbo DNase (Ambion) for 60 min at 37 °C. Cells (2×10^5) were infected with 2×10^4 RT units of VSV G-pseudotyped HIV-1-GFP or control HIV-1-GFP lacking envelope glycoproteins, and genomic DNA was isolated at various time points (0–48 h). We quantified early HIV-1 reverse transcripts with primers ert2f and ert2r and the ERT2 probe⁸, and late HIV-1 reverse transcripts with primers MH531 and MH532 and the probe LRT-P³⁰ as described. Reaction mixtures contained Taqman universal master mix (PE Biosystems), 300 nM primers, 100 nM probe and 500 ng of genomic DNA. The PCR conditions were 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C, on an ABI Prism 7700 (Applied Biosystems).

Cloning of TRIM5 isoforms and TRIM5 α_{rh} mutagenesis

The human TRIM5 α open reading frame (ORF) was amplified from a kidney cDNA library (Clontech), using primers derived from National Center for Biotechnology Information (NCBI) RefSeq NM_033034 and inserted into pLPCX (Clontech) to create pTRIM5 α_{hu} . The predicted sequence of this TRIM5 α_{hu} differs in three amino acid residues from NCBI RefSeq NM_033034, which is derived from the TRIM5 α of the retinoic-acid-induced NT2 neuronal precursor line. We amplified the rhesus monkey TRIM5 γ ORF from the PRL cDNA library with primers derived from NCBI RefSeq NM_033092 and inserted it into pLPCX to create pTRIM5 γ_{rh} . In the pTRIM5 α_{rh} -HA and pTRIM5 α_{hu} -HA plasmids, an in-frame sequence encoding the influenza virus HA epitope tag was included at the 3' end of the TRIM5 α_{rh} and TRIM5 α_{hu} sequences, respectively. The pTRIM5 γ_{rh} -HA plasmid contains a sequence encoding the initiator methionine and the HA epitope tag at the 5' end of the TRIM5 γ_{rh} sequence. Cysteine to alanine changes (C15A and C15A/C18A) were introduced into the TRIM5 α_{rh} RING domain by using the Quick-change mutagenesis kit (Stratagene).

Immunoblotting

HA-tagged proteins were expressed in HeLa cells by transfection with Lipofectamine 2000

(Invitrogen) or by transduction with pLPCX vectors as described above. The HA-tagged proteins were detected in whole-cell lysates (100 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 7.5), 10% glycerol, 1% Nonidet P40) by western blotting with horseradish peroxidase (HRP)-conjugated 3F10 antibody (Roche). β -Actin was detected with A5441 antibody (Sigma).

RNA interference

Eight siRNAs directed against TRIM5 α were designed using the siRNA Selection Program (Whitehead Institute for Biomedical Research, 2003) and purchased from Dharmacon RNA Technologies: siRNA 1, 5'-GCUCAGGGAGGCAAGUUGdTdT-3'; siRNA 2, 5'-GAGAAAGCUUCCUGGAAGdTdT-3'; siRNA 3, 5'-GCCUACGAA GUCUGAAACdTdT-3'; siRNA 4, 5'-GGAGAGUGUUCGAGCUCdTdT-3'; siRNA 5, 5'-CCUUCUACACACUCAGCCdTdT-3'; siRNA 6, 5'-CGUCCUGCACUCAU CAGUGdTdT-3'; siRNA 7, 5'-CAGCCUUCUUAUUAUCGdTdT-3'; siRNA 8, 5'-UCCUGUCUCUCCAUCAUGACdTdT-3'.

Four of the siRNAs (1–4) were directed against TRIM5 coding sequences common to the messenger RNAs of all of the TRIM5 isoforms. Four of the siRNAs (5–8) were directed against the 3' UTR specific to the mRNAs encoding TRIM5 α and TRIM5 ϵ . Two of the siRNAs (2 and 8) showed some toxicity to HeLa cells and were not studied further. After transfection into PRL cells, the remaining six siRNAs all showed ability to relieve the block to HIV-1-GFP infection (Fig. 4 and data not shown). A control siRNA (Nonspecific control duplex 1, 5'-AUGAACGUGAAUUGCUCAAUU-3'; Dharmacon RNA Technologies) was included in the experiments. HeLa cells (1×10^5) or PRL cells (5×10^4) were seeded in six-well plates and transfected with 120 nM siRNA by using 10 μ l of Oligofectamine (Invitrogen). After 48 h, cells were reseeded for HIV-1-GFP infection. In some experiments, HeLa and PRL cells were transduced with a pLPCX vector encoding TRIM5 α -escape or an empty pLPCX vector as a control. After 48 h, the cells were transfected with siRNA; and after 2 d, the cells were replated and used for infection. In the vector encoding TRIM5 α -escape, silent mutations at the siRNA 3 recognition site changed the wild-type sequence from 5'-GCCTACGAAGTCTGAAAC-3' to 5'-GGTTAACGAAGAGCGAAAC-3'.

Received 19 November 2003; accepted 13 January 2004; doi:10.1038/nature02343.

- LaBonte, J., Babcock, G., Patel, T. & Sodroski, J. Blockade of human immunodeficiency virus (HIV-1) of New World monkey cells occurs primarily at the stage of virus entry. *J. Exp. Med.* **196**, 431–435 (2002).
- Shibata, R., Sakai, H., Kawamura, M., Tokunaga, K. & Adachi, A. Early replication block of human immunodeficiency virus type 1 in monkey cells. *J. Gen. Virol.* **76**, 2723–2730 (1995).
- Himathongkham, S. & Luciw, P. A. Restriction of HIV-1 (subtype B) replication at the entry step in rhesus macaque cells. *Virology* **219**, 485–488 (1996).
- Hofmann, W. *et al.* Species-specific, postentry barriers to primate immunodeficiency virus infection. *J. Virol.* **73**, 10020–10028 (1999).
- Owens, C. M., Yang, P. C., Göttinger, H. & Sodroski, J. Human and simian immunodeficiency virus capsid proteins are major viral determinants of early, post-entry replication blocks in simian cells. *J. Virol.* **77**, 726–731 (2003).
- Kootstra, N. A., Munk, C., Tonnau, N., Landau, N. R. & Verma, I. M. Abrogation of postentry restriction of HIV-1 based lentiviral vector transduction in simian cells. *Proc. Natl Acad. Sci. USA* **100**, 1298–1303 (2003).
- Cowan, S. *et al.* Cellular inhibitors with Fv-1-like activity restrict human and simian immunodeficiency virus tropism. *Proc. Natl Acad. Sci. USA* **99**, 11914–11919 (2002).
- Besnier, C., Takeuchi, Y. & Towers, G. Restriction of lentivirus in monkeys. *Proc. Natl Acad. Sci. USA* **99**, 11920–11925 (2002).
- Munk, C., Brandt, S. M., Lucero, G. & Landau, N. R. A dominant block to HIV-1 replication at reverse transcription in simian cells. *Proc. Natl Acad. Sci. USA* **99**, 13843–13848 (2002).
- Kanki, P. J., Alroy, J. & Essex, M. Isolation of T-lymphotropic retrovirus related to HTLV-III/LAV from wild-caught African green monkeys. *Science* **230**, 951–954 (1985).
- Reddy, B. A., Etkin, L. D. & Freemont, P. S. A novel zinc finger coiled-coil domain in a family of nuclear proteins. *Trends Biochem. Sci.* **17**, 344–345 (1992).
- Borden, K. L. RING fingers and B-boxes: zinc-binding protein–protein interaction domains. *Biochem. Cell Biol.* **76**, 351–358 (1998).
- Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO J.* **20**, 2140–2151 (2001).
- Xu, L. *et al.* BTBD1 and BTBD2 colocalize to cytoplasmic bodies with the RBCC/tripartite motif protein, TRIM5 ϵ . *Exp. Cell Res.* **288**, 84–93 (2003).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry co-factor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
- Li, J., Lord, C. I., Haseltine, W., Letvin, N. L. & Sodroski, J. Infection of cynomolgus monkeys with a chimeric HIV-1/SIV_{mac} virus that expresses the HIV-1 envelope glycoproteins. *J. AIDS* **5**, 639–646 (1992).
- Dorfman, T. & Göttinger, H. G. The human immunodeficiency virus type 1 capsid p2 domain confers sensitivity to the cyclophilin-binding drug SDZ NIM 811. *J. Virol.* **70**, 5751–5757 (1996).
- Waterman, H., Levkowitz, G., Alroy, I. & Yarden, Y. The RING finger of c-CBL mediates desensitization of the epidermal growth factor receptor. *J. Biol. Chem.* **274**, 22151–22154 (1999).
- Owens, C. M. *et al.* Binding and susceptibility to post-entry restriction factors in monkey cells are specified by distinct regions of the human immunodeficiency virus type 1 capsid. *J. Virol.* (in the press).
- Fassati, A. & Goff, S. P. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J. Virol.* **75**, 3626–3635 (2001).
- Forshey, B. M., von Schwedler, U., Sundquist, W. I. & Aiken, C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* **76**, 5667–5677 (2002).
- Schwartz, O., Maréchal, V., Friguet, B., Arenzana-Seisdedos, F. & Heard, J.-M. Antiviral activity of the proteasome on incoming human immunodeficiency virus type 1. *J. Virol.* **72**, 3845–3850 (1998).

- Turelli, P. *et al.* Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol. Cell* **7**, 1245–1254 (2001).
- Marcello, A. *et al.* Recruitment of human cyclin T1 to nuclear bodies through direct interaction with the PML protein. *EMBO J.* **22**, 2156–2166 (2003).
- Bonilla, W. V. *et al.* Effects of promyelocytic leukemia protein on virus–host balance. *J. Virol.* **76**, 3810–3818 (2002).
- Chee, A. V., Lopez, P., Pandolfi, P. P. & Roizman, B. Promyelocytic leukemia protein mediates interferon-based anti-herpes simplex virus 1 effects. *J. Virol.* **77**, 7101–7105 (2003).
- Rhodes, D. A. *et al.* The 52,000 MW Ro/SS-A autoantigen in Sjogren's syndrome/systemic lupus erythematosus (Ro52) is an interferon- γ inducible tripartite motif protein associated with membrane proximal structures. *Immunology* **106**, 246–256 (2002).
- Toniato, E. *et al.* TRIM8/GERP RING finger protein interacts with SOCS-1. *J. Biol. Chem.* **277**, 37315–37322 (2002).
- Mariani, R. *et al.* Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**, 21–31 (2003).
- Butler, S., Hansen, M. S. T. & Bushman, F. A quantitative assay for HIV DNA integration *in vivo*. *Nature Med.* **7**, 631–634 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Farnum and Y. McLaughlin for manuscript preparation; and S. Basmaciogullari, R. Lu, N. Vandegraaff and A. Engelman for advice and reagents. This work was supported by grants from the NIH, by a Center for AIDS Research Award, by the International AIDS Vaccine Initiative, and by the Bristol-Myers Squibb Foundation. M.S. is supported by a National Defense Science and Engineering Fellowship and is a Fellow of the Ryan Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S. (joseph_sodroski@dfci.harvard.edu).

The large-conductance Ca²⁺-activated K⁺ channel is essential for innate immunity

Jatinder Ahluwalia¹, Andrew Tinker¹, Lucie H. Clapp¹, Michael R. Duchon², Andrey Y. Abramov², Simon Pope¹, Muriel Nobles¹ & Anthony W. Segal¹

¹Department of Medicine and ²Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

Neutrophil leukocytes have a pivotal function in innate immunity. Dogma dictates that the lethal blow is delivered to microbes by reactive oxygen species (ROS) and halogens^{1,2}, products of the NADPH oxidase, whose impairment causes immunodeficiency. However, recent evidence indicates that the microbes might be killed by proteases, activated by the oxidase through the generation of a hypertonic, K⁺-rich and alkaline environment in the phagocytic vacuole³. Here we show that K⁺ crosses the membrane through large-conductance Ca²⁺-activated K⁺ (BK_{Ca}) channels. Specific inhibitors of these channels, iberiotoxin and paxilline, blocked oxidase-induced ⁸⁶Rb⁺ fluxes and alkalization of the phagocytic vacuole, whereas NS1619, a BK_{Ca} channel opener, enhanced both. Characteristic outwardly rectifying K⁺ currents, reversibly inhibited by iberiotoxin, were demonstrated in neutrophils and eosinophils and the expression of the α -subunit of the BK channel was confirmed by western blotting. The channels were opened by the combination of membrane depolarization and elevated Ca²⁺ concentration, both consequences of oxidase activity. Remarkably, microbial killing and digestion were abolished when the BK_{Ca} channel was blocked, revealing an essential and unexpected function for this K⁺ channel in the microbicidal process.

The NADPH oxidase is required for normal immunity^{1,2}; where defective it results in chronic granulomatous disease (CGD)⁴. The

oxidase transfers electrons from NADPH to oxygen, forming superoxide anions (O_2^-) in the phagocytic vacuole⁴. This process is electrogenic⁵. The charge generated by the passage of electrons across the membrane is compensated for⁶ in part through a K^+ flux³, which seems essential for microbial killing by these cells³. Because of the importance and novelty of this process, we have now sought the identity of the K^+ channel involved.

The NADPH oxidase pumps O_2^- into the vacuole, which, together with its dismutation product O_2^{2-} , becomes protonated, consuming H^+ ions. The observed increase in vacuolar pH (ref. 7) depends on charge compensation with ions other than protons from the cytoplasm. We predicted that blockade of the channel through which K^+ enters the vacuole would cause pH to decrease, whereas opening the channels excessively would cause pH to increase above normal. Using fluorescein conjugated to *Staphylococcus aureus* as a pH indicator we found that modulators of the BK_{Ca} channel produced the appropriate alterations (Fig. 1a). Iberitoxin (for which the concentration for 50% inhibition (IC_{50}) is 9.7 nM) and paxilline (IC_{50} 17 nM), both highly selective and potent inhibitors^{8,9}, prevented alkalinization (Fig. 1a, b), as did the oxidase inhibitor diphenylene iodonium (DPI)⁵, whereas other K^+ channel blockers—4-aminopyridine (4-AP)¹⁰, apamin¹¹, glibenclamide¹² and anandamide¹³—did not (Fig. 1a, b). The selective opener NS1619 (ref. 14) elevated pH above normal (Fig. 1a, b), unlike the K_{ATP} channel opener levcromakalim¹⁴ (Fig. 1a).

$^{86}Rb^+$ is commonly used as a surrogate for K^+ in flux studies. When the oxidase is activated by 12-O-tetradecanoylphorbol-13-

acetate (TPA), the O_2^- and $^{86}Rb^+$ are expelled into the extracellular medium. Figure 1c shows that the $^{86}Rb^+$ flux increased fourfold after stimulation with TPA; an efflux approaching this was also induced by opening the BK_{Ca} channel with NS1619 and was even further enhanced by combining this opener with TPA. The K^+ efflux that resulted from stimulation with TPA was completely abrogated by iberitoxin or paxilline, confirming that the efflux of $^{86}Rb^+$ occurred through BK_{Ca} channels. The requirement for an active oxidase was shown by the inhibition of $^{86}Rb^+$ flux by DPI. The release of the isotope induced by NS1619 was also completely abolished by iberitoxin. Once again, 4-AP was without effect. Similar results were obtained with eosinophils (Fig. 1d).

The expression of the BK_{Ca} channels was detected in cell membranes and in membrane obtained from cytoplasmic granules (Fig. 1e), but not in the cytoplasm of neutrophils, by western blotting with an antibody to the α -subunit of the channel. Eosinophils contained about one-half of the amount of protein present in neutrophils (compare lanes 1 and 2 in Fig. 1e). Reverse-transcriptase-mediated polymerase chain reaction (RT-PCR) on mRNA derived from HL-60 cells with primers corresponding to the sequence of the α -subunit of the BK_{Ca} channel produced the predicted 479-base-pair product in cells induced with dimethyl sulphoxide (DMSO) to differentiate along the granulocyte lineage, but not in undifferentiated cells. The sequence of this PCR product corresponded exactly to nucleotides 2,822–3,301 of the *hSlo* complete coding sequence (U11717)¹⁵.

In patch-clamp studies¹⁶ of neutrophils, we observed small outward currents averaging about 250 pA at +140 mV under resting conditions. Current density varied from cell to cell; this might reflect variable activation of the oxidase by contact with the glass of the coverslip or pipette. After the addition of TPA, a large outwardly rectifying current developed at potentials positive to -30 mV, taking several minutes to develop, and tending to increase slightly with time thereafter. Small, variable inward currents, and inward tails, were observed at hyperpolarized potentials of less than -100 mV, and these became more obvious as the pulse times were increased. All these currents were completely and reversibly inhibited by iberitoxin (Fig. 2a, upper panels, and b).

The accepted model for such electrophysiological studies is the eosinophil^{16,17}, and similar $^{86}Rb^+$ fluxes (Fig. 1d) and TPA-activated outward currents, which were reversibly inhibited by iberitoxin (Fig. 2c, top panels, and d), were also seen in these cells. Because the currents measured in eosinophils were only about one-third of those observed in neutrophils, this might suggest a lower oxidase activity in these cells, despite reports to the contrary¹⁸. We therefore measured the rates of both oxygen consumption and superoxide generation by these two cell types and showed both to be two to three times as rapid in neutrophils as in eosinophils under our experimental conditions (Fig. 2g).

In addition we performed cell-attached single-channel recordings from TPA-stimulated neutrophils and observed channels activated by depolarization with a characteristically large single-channel conductance¹⁵ (Fig. 2e).

It has been suggested that the charge induced by electron translocation through the NADPH oxidase (I_e) is compensated for by proton efflux^{5,6,16}. One of the foundations for this assumption is that both Zn^{2+} and Cd^{2+} , known proton channel blockers^{16,19}, were also thought to inhibit O_2^- production^{5,16}. We examined the effect of Zn^{2+} and Cd^{2+} on $^{86}Rb^+$ release, which we expected to be depressed through an inhibition of I_e . Perversely, we found that $^{86}Rb^+$ release was greater in TPA-stimulated neutrophils and eosinophils in their presence (Fig. 2f) and that Zn^{2+} , at concentrations three orders of magnitude greater than those causing almost complete blockage of proton channels²⁰, was also without effect on the currents from neutrophils (Fig. 2a, lower panels) and eosinophils (Fig. 2c, middle panels, and d). We therefore re-examined the influence of Zn^{2+} and Cd^{2+} on oxidase activity.

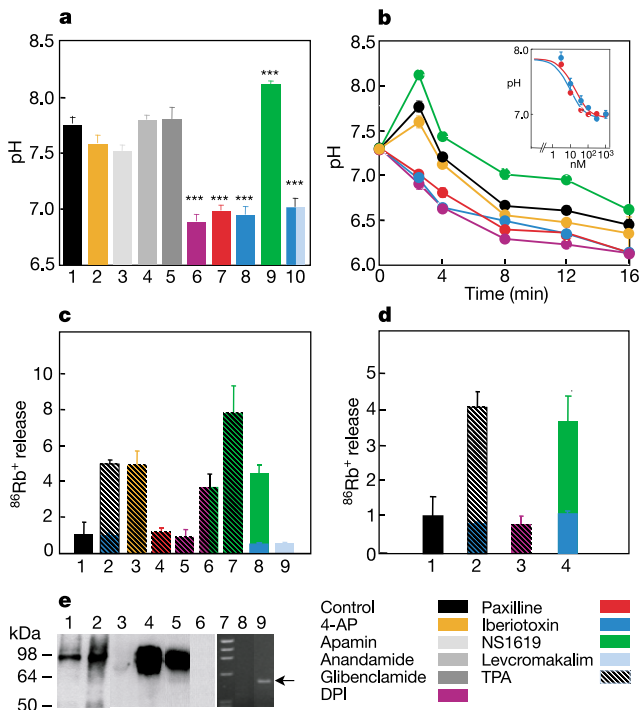


Figure 1 BK_{Ca} channels influence the pH within the phagocytic vacuole and $^{86}Rb^+$ efflux from neutrophils and eosinophils. **a**, Vacuolar pH at 150 s (means $\pm$ s.e.m.; three asterisks, $P < 0.001$ compared with control). **b**, Time course of pH changes. Inset, dose response of depression of pH by iberitoxin and paxilline. **c**, **d**, $^{86}Rb^+$ release from neutrophils (**c**) and from eosinophils (**d**). For control cells the release was $14 \pm 3\%$ and was normalized to 1. **e**, Western blots for α -subunit of BK_{Ca} channel. Lane 1, eosinophil homogenate; lane 2, neutrophil homogenate; lane 3, cytosol; lane 4, granule membranes; lane 5, plasma membrane; lane 6, plasma membrane plus immunizing peptide. PCR fragment (arrow) for *hSlo* product from differentiated (lane 9) and undifferentiated (lane 8) HL-60 cells. Lane 7, markers.

Whereas oxygen consumption, the true measure of NADPH oxidase activity, was unaffected, both Zn^{2+} and Cd^{2+} almost completely inhibited the reduction of ferricytochrome *c* by O_2^- (Fig. 2g) by accelerating the dismutation of O_2^- to H_2O_2 . In a system in which xanthine–xanthine-oxidase generated O_2^- , 3 mM concentrations of these elements induced the dismutation of O_2^- to H_2O_2 at a rate indistinguishable from that catalysed by $1 \mu\text{g ml}^{-1}$ superoxide dismutase (Fig. 2h).

BK_{Ca} channels are classically opened by the combination of membrane depolarization and elevated cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]_{\text{c}}$; ref. 21). The same was found to hold true for this channel in neutrophils (Fig. 3a) and eosinophils (Fig. 3b). Neither depolarizing the membrane by elevating external $[\text{K}^+]$ to 136 mM nor elevating the $[\text{Ca}^{2+}]_{\text{c}}$ with the ionophore A23187, which produced a large fluorescence signal from the intracellular Ca^{2+} indicator Fluo-4 acetoxymethyl ester (AM) (Fig. 3g), was sufficient to open the K^+ channel fully (Fig. 3a). $^{86}\text{Rb}^+$ release was slightly increased by the addition of A23187, which is a weak stimulus for NADPH oxidase and membrane depolarization²². The combination of membrane depolarization and A23187 caused as much $^{86}\text{Rb}^+$ efflux from both neutrophils and eosinophils (Fig. 3a, b) as did stimulation with TPA, and in both cell types this was totally blocked by iberiotoxin. Chelation of $[\text{Ca}^{2+}]_{\text{c}}$ with bis-(*o*-aminophenoxy)ethane N,N,N',N' -tetra-acetic acid (BAPTA) AM and extracellular EGTA also completely blocked $^{86}\text{Rb}^+$ release from TPA-stimulated cells (Fig. 3a) without inhibiting oxygen

consumption (not shown), strongly suggesting that TPA must raise $[\text{Ca}^{2+}]_{\text{c}}$.

Although TPA stimulation is well known to depolarize the neutrophil plasma membrane²³, it is generally thought not to elevate $[\text{Ca}^{2+}]_{\text{c}}$ (ref. 22). We therefore measured changes in $[\text{Ca}^{2+}]_{\text{c}}$ after exposure to TPA. Using Fluo-4 to study $[\text{Ca}^{2+}]_{\text{c}}$ in cell populations (Fig. 3g), and fura-2 and indo-1 in fluorescence imaging at the level of the single cell²⁴ (Fig. 3c, e), we found that an increase in $[\text{Ca}^{2+}]_{\text{c}}$ coincided with initiation of the oxidase activity, measured with hydroethidine²⁵ (Fig. 3c). The increase in $[\text{Ca}^{2+}]_{\text{c}}$ was inhibited by DPI (Fig. 3d), linking the signalling pathway to the activity of the oxidase.

Increases in intracellular $[\text{Ca}^{2+}]$ were not blocked by chelating external Ca^{2+} with EGTA but were inhibited by pretreatment with the microsomal Ca^{2+} -ATPase inhibitor thapsigargin²⁶ (Fig. 3g), which depletes intracellular Ca^{2+} stores, indicating an intracellular source for the Ca^{2+} . This posed the question of how this Ca^{2+} is mobilized by oxidase activity. One possibility seemed to be through a decrease in pH, because activation of the NADPH oxidase acidifies the cytosol²⁷. In cells labelled with Fluo-4 AM, elevated $[\text{Ca}^{2+}]_{\text{c}}$ was prevented by pretreatment with the protonophore carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), and acidification of the cytosol after an acid loading protocol with NH_4Cl (ref. 28) resulted in a rapid increase in $[\text{Ca}^{2+}]_{\text{c}}$. The oxidase pumps 4 mol l^{-1} of electrons into the vacuole³, and charge separation across the membrane leaves relatively large concentrations of protons on the cytosolic side. This

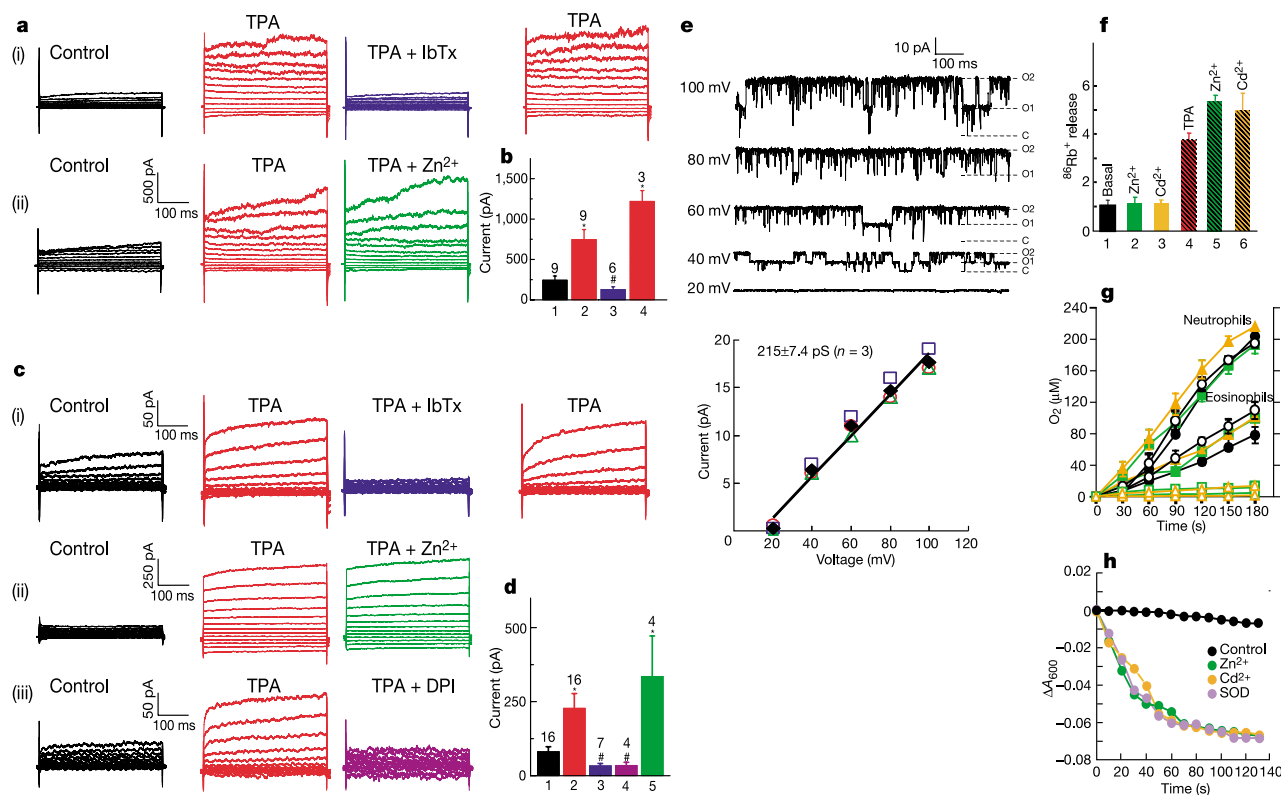


Figure 2 BK_{Ca} currents in granulocytes. **a**, Representative recordings from two neutrophils showing activation of outward currents by TPA and subsequent reversible inhibition by iberiotoxin (IbTx; upper panels) but not by Zn^{2+} (lower panels) in the presence of TPA. **b**, Bar graph of pooled data (lane 1, control; lane 2, TPA; lane 3, TPA + iberiotoxin; lane 4, iberiotoxin washout + TPA). Asterisk, $P < 0.01$; hash sign, $P \leq 0.05$ compared with control. Numbers of samples are shown above the columns. **c**, Representative recordings from three different eosinophils showing activation of outward currents by TPA that were reversibly inhibited by iberiotoxin (top panels) and insensitive to block by Zn^{2+} (middle panels) but were inhibited by DPI (bottom panels).

d, Pooled data. Lane 1, control; lane 2, TPA; lane 3, iberiotoxin; lane 4, DPI; lane 5, Zn^{2+} . Numbers of samples are shown above the columns. **e**, Characteristic cell-attached single-channel recordings from a TPA-stimulated neutrophil. A single-channel current–voltage relationship ($n = 3$) is shown below. **f**, Neither 3 mM Zn^{2+} nor Cd^{2+} inhibited the TPA-stimulated (hatched columns) release of $^{86}\text{Rb}^+$. **g**, Zn^{2+} (squares) and Cd^{2+} (triangles) inhibit cytochrome *c* reduction (open symbols) but not oxygen consumption (filled symbols) by TPA-stimulated (circles) neutrophils and eosinophils. **h**, Zn^{2+} and Cd^{2+} dismutate O_2^- at the same rate as superoxide dismutase (SOD; $1 \mu\text{g ml}^{-1}$).

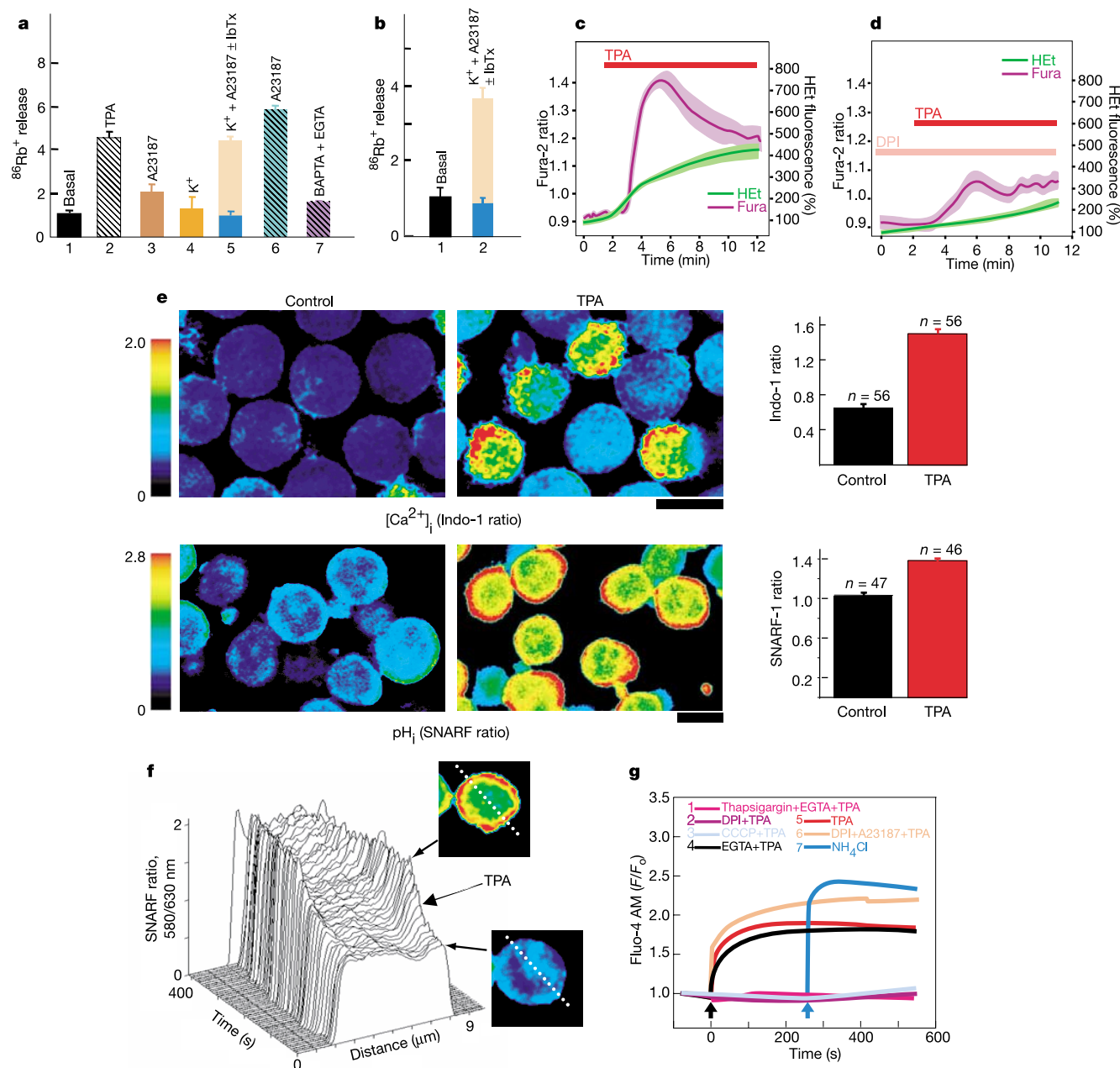


Figure 3 Ca^{2+} and membrane depolarization open BK_{Ca} channels. **a**, **b**, Effects of TPA (hatched) in the presence or absence of other agents on $^{86}\text{Rb}^+$ release from neutrophils (**a**) or eosinophils (**b**). IbTx, iberiotoxin. **c**, **d**, Traces of cells loaded with fura-2 and hydroethidine (HET). TPA increased $[\text{Ca}^{2+}]_c$ (fura-2) and increased ROS generation (HET) (means $\pm$ s.d. from 60 cells) (**c**), both of which were suppressed by DPI (**d**). **e**, **f**, Indo-1 as the $[\text{Ca}^{2+}]_i$ indicator (upper panels in **e**) showed a significant increase in $[\text{Ca}^{2+}]_c$, which appeared more pronounced under the plasmalemma after TPA (means $\pm$ s.e.m.). TPA

decreased pH_i (imaged with carboxy-SNARF-1; lower panels in **e**), particularly under the plasmalemma, as shown with images of a single neutrophil (**f**). **g**, FLIPR measurements of $[\text{Ca}^{2+}]_c$ with Fluo-4 AM. TPA induced a rise in $[\text{Ca}^{2+}]_c$ (curve 5) that was blocked by DPI (curve 2), CCCP (curve 3) and thapsigargin and EGTA (curve 1) but not by EGTA alone (curve 4). Curve 6 shows the effect of TPA, DPI and A23187; curve 7 shows the effect of acidification of cytosol with NH_4Cl washout (blue arrow). (Agents were applied at the black arrow.)

was detected by confocal imaging of pH_i with the dual-wavelength ratiometric indicator 5'-(and 6')-carboxy-seminaphthorhodofluor-1 (carboxy-SNARF), which revealed that TPA caused an intracellular acidification clearly localized beneath the plasma membrane (Fig. 3e, f). In addition, confocal imaging under ultraviolet with Indo-1 as a ratiometric dual-emission Ca^{2+} indicator also showed that the increase in $[\text{Ca}^{2+}]_c$ was also predominantly localized under the plasma membrane in many cells, although in others it was more generalized.

Specific and complete inhibitors of the BK_{Ca} channel made it

possible to selectively examine the role of K^+ flux through BK_{Ca} channels, and the consequent elevation of vacuolar pH, in the killing of microbes³. Paxilline and iberiotoxin, but not 4-AP, completely inhibited the killing of *S. aureus* and *Serratia marcescens*—which classically infect CGD patients⁴—and *Candida albicans* (Fig. 4a–c). This failure of killing occurred despite totally normal oxidase activity, phagocytosis and iodination (Fig. 4e–g) and a normal composition of granule enzymes. The disruptive effects of channel block on the enzymatic digestion of dead bacteria were also striking (Fig. 4d).

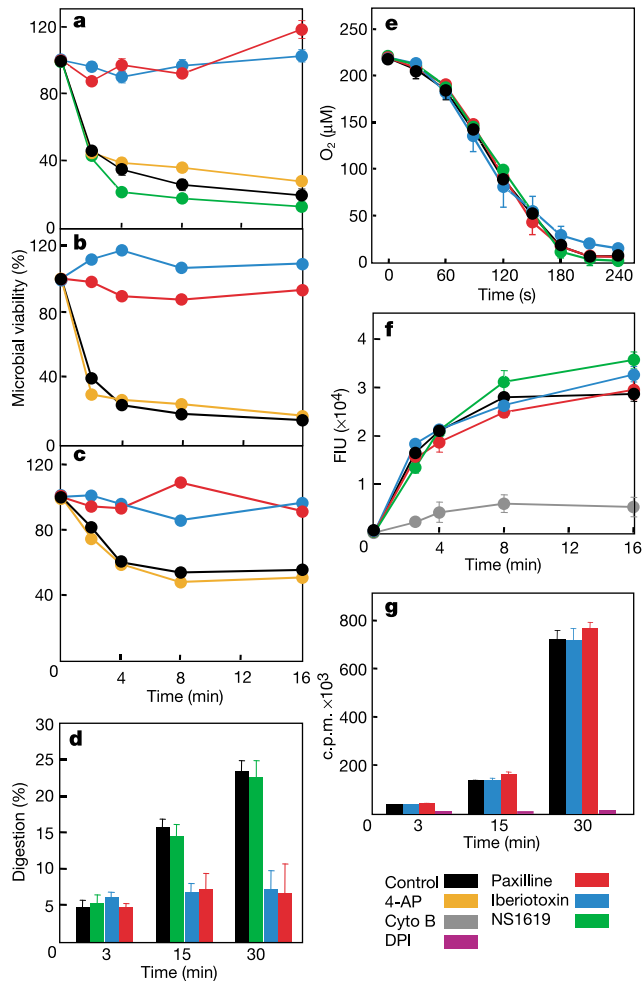


Figure 4 Inhibitors of BK_{Ca} channels abolish microbial killing and digestion. **a–c**, Killing of *S. aureus* (**a**; $n = 5$), *S. marcescens* (**b**; $n = 2$) and *C. albicans* (**c**; $n = 2$) was inhibited by iberiotoxin and paxilline, whereas survival in the presence of 4-AP or NS1619 was no different from that of control neutrophils (circles). **d**, Digestion of ^{35}S -methionine-labelled *S. aureus* was prevented by both iberiotoxin ($n = 3$) and paxilline ($n = 3$) but was similar to that of the control ($n = 3$) in the presence of NS1619 ($n = 3$). **e–g**, None of these modulators of the BK_{Ca} channel altered oxygen consumption (**e**), phagocytosis (**f**; as fluorescent intensity units (FIU), $n = 3$, impaired by cytochalasin B (cyto B) or iodination (**g**; inhibited by DPI).

These data have significance beyond the inherent value of defining the precise molecular mechanisms involved in a physiological process of paramount importance to survival. The perception that neutrophils kill microbes through toxic oxygen radicals and their metabolites provided much of the biological basis for the theories relating the toxicity of oxygen radicals to the pathogenesis of a wide variety of human diseases, and the development of antioxidant drugs for their treatment²⁹. These theories and treatments merit re-evaluation. □

Methods

Concentrations of reagents

Unless otherwise stated, the concentrations of reagents were as follows: A23187, 10 μ M; apamin, 200 nM; anandamide, 3 μ M; 4-AP, 4 mM; Cd^{2+} , 3 mM; DPI, 5 μ M; EGTA, 3 mM; glibenclamide, 10 μ M; iberiotoxin, 100 nM; levcromakalim, 10 μ M; NS1619, 30 μ M; paxilline, 300 nM; TPA, 1 μ g ml⁻¹; Zn^{2+} , 3 mM; superoxide dismutase, 1 μ g ml⁻¹.

Purification of cells

Granulocytes were purified from human blood. Eosinophils were isolated by negative selection with anti-CD16 immunomagnetic beads¹⁶.

Measurement of vacuolar pH and $[Ca^{2+}]_i$

Vacuolar pH was determined from the excitation spectrum of fluorescein-labelled, IgG-coated *S. aureus*⁷. Neutrophils, purified from human blood in Dulbecco's phosphate-buffered saline (PBS; 140 mM NaCl, 10 mM KCl, 10 mM $Na_2H_2PO_4$, 5 mM glucose pH 7.3) were plated into black-walled clear-based 96-well plates at a density of 1.5×10^7 per well, and mixed with inhibitors and/or openers for 3 min at 37 °C. Fluorescein-labelled, IgG-coated *S. aureus* cells (2.5×10^6) were then mixed with the neutrophils and the fluorescence (λ_{ex} 494 nm, λ_{em} 518 nm) from each well was measured at intervals in a fluorimetric imaging plate reader (FLIPR; Molecular Devices). A comparison of fluorescence was made with a calibration curve generated with free bacteria in buffers of different pH values. Curves in Fig. 1b were fitted by nonlinear regression analysis using the equation for a sigmoid concentration–response curve (GraphPad Prism, GraphPad) with the Hill coefficient constrained to one.

For measurements of $[Ca^{2+}]_i$, cells were plated on black-walled clear-based 96-well plates at a density of 10^6 per well, incubated with Dulbecco's PBS containing Fluo-4 AM (4 μ M; Molecular Probes) for 2 h and washed three times to remove extracellular Fluo-4 AM. Fluorescence (λ_{ex} 494 nm, λ_{em} 518 nm) from each well was measured before and after the addition of DPI and/or TPA and/or A23187. In thapsigargin experiments, neutrophils were preincubated for 30 min in PBS containing 500 nM thapsigargin and 3 mM EGTA, after which Fluo-4 AM was added. Washing steps were performed in Ca^{2+} -free Dulbecco's PBS supplemented with 3 mM EGTA. In other experiments, neutrophils were preincubated in Ca^{2+} -free Dulbecco's PBS containing 3 mM EGTA for 10 min before the addition of TPA. The effect of intracellular pH changes on $[Ca^{2+}]_i$ was studied with NH_4Cl as described²⁸.

Measurement of $^{86}Rb^+$ fluxes

$^{86}Rb^+$ efflux was measured as described previously³. Labelled cells were incubated with iberiotoxin, paxilline or 4-AP for 3 min before stimulation with TPA. The calcium ionophore A23187, with BAPTA, EGTA and the channel openers NS1619 or levcromakalim, were added with or without TPA unless otherwise indicated. For depolarization studies, neutrophils or eosinophils were incubated in HEPES buffer containing 136 mM KCl.

Electrophysiology

Patch-clamp studies were performed on freshly isolated neutrophils and eosinophils. Membrane currents were recorded with an Axopatch 200B amplifier (Axon Instruments). Data were acquired and analysed with a Digidata interface (1200A or 1322; Axon Instruments) and pClamp software (version 6.0 and 8.0; Axon Instruments). Cells were voltage-clamped at -30 mV and pulsed for 400 ms from -100 mV to $+140$ mV in 20-mV increments. Currents were measured at the end of the pulse at $+140$ mV. Statistical analysis was performed by using one-way analysis of variance with a Bonferroni or Student–Newman–Keuls post hoc test. For the perforated-patch experiments, a stock solution of 100 mg ml⁻¹ amphotericin B in DMSO was prepared and diluted in the pipette solution to give a final concentration of 200 μ g ml⁻¹ amphotericin B. Stable access was obtained after 20 min, and cell capacitance was about 1–5 pF. Series resistance was not routinely corrected for. Agents were applied to the bath with a gravity-driven perfusion system. The pipette solution contained (in mM): 140 KCl, 10 NaCl, 2 MgCl₂, 0.7 CaCl₂, 1 EGTA and 10 HEPES (made to pH 7.3 with KOH). Cells were perfused with an extracellular buffer containing (in mM): 140 NaCl, 2.5 KCl, 0.5 MgCl₂, 1.2 CaCl₂, 101 HEPES and 5 glucose (made to pH 7.4 with NaOH).

Confocal imaging

Cells were loaded with either 5 mM Indo-1 or 5 mM carboxy-SNARF AM for 20 min, washed in Hanks balanced salt solution and plated at low density on 22-mm glass coverslips, which were mounted on the stage of a Zeiss ultraviolet–visible confocal microscope equipped with a META detection system. Indo-1 fluorescence was excited at 351 nm and an emission ratio was acquired on two channels at 410 and 490 nm, and SNARF emission was detected at 530 and 580 nm, each with a 25-nm bandwidth. Indo-1 tends to bleach, so laser power was kept as low as possible; images were therefore a little noisy. The dual-emission ratiometric measurements are important here to avoid artefacts due to cell movement after stimulation. Oxidase activation or effective inhibition was confirmed with hydroethidine (1 mM) to monitor radical formation, excited at 543 nm and measured at 580–650 nm. Drugs were added to the chamber as concentrated stocks to reach the desired final concentration.

In addition, images of cells loaded with fura-2 AM (5 mM for 20 min) were acquired at low power on a cooled charge-coupled device-based imaging system. Fluorescence was excited alternately at 340 and 380 nm and measured at more than 510 nm.

Subcellular fractionation of neutrophils, western blotting and PCR

Plasma and granules were prepared from neutrophil post-nuclear supernatants, and granule membranes removed as described³. Membranes were purified at the interface of discontinuous sucrose gradients of 15% on 30% sucrose (w/w) western blots, which were processed as described¹⁵. Protein run on gel was as follows: homogenates of eosinophils and neutrophils, 18 μ g; granule and plasma membranes, 10 μ g; cytosol, 20 μ g. In control experiments, the antibody was incubated with the immunizing peptide (50 μ g ml⁻¹) for 30 min before use. RT–PCR of BK_{Ca} mRNA extracted from HL-60 cells before and after 5 days of DMSO-induced differentiation was as described¹⁵. The marker used was ΦX -174 Hae.

Measurements of neutrophil function *in vitro*

Microbial killing, digestion, iodination, oxygen consumption and superoxide generation were studied as described³. For killing studies, 10^8 neutrophils in 1 ml Dulbecco's PBS were

preincubated with iberiotoxin, paxilline, 4-AP or NS1619 for 3 min in a rapidly stirred oxygenated chamber at 37 °C, then mixed with 10⁸ IgG-opsonized microbes. Phagocytosis was measured as described⁷ using fluorescent rather than radiolabelled bacteria.

Superoxide dismutase activity

H₂O₂ production was measured in a xanthine–xanthine-oxidase O₂²⁻-generating system by the bleaching of phenol red³⁰.

Received 18 July 2003; accepted 20 January 2004; doi:10.1038/nature02356.

- Klebanoff, S. J. Antimicrobial mechanisms in neutrophilic polymorphonuclear leukocytes. *Semin. Hematol.* **12**, 117–142 (1975).
- Babior, B. M. Phagocytes and oxidative stress. *Am. J. Med.* **109**, 33–44 (2000).
- Reeves, E. P. et al. Killing activity of neutrophils is mediated through activation of proteases by K⁺ flux. *Nature* **416**, 291–297 (2002).
- Thrasher, A. J., Keep, N. H., Wientjes, F. & Segal, A. W. Chronic granulomatous disease. *Biochim. Biophys. Acta* **1227**, 1–24 (1994).
- Henderson, L. M., Chappell, J. B. & Jones, O. T. The superoxide-generating NADPH oxidase of human neutrophils is electrogenic and associated with an H⁺ channel. *Biochem. J.* **246**, 325–329 (1987).
- Henderson, L. M., Chappell, J. B. & Jones, O. T. Superoxide generation by the electrogenic NADPH oxidase of human neutrophils is limited by the movement of a compensating charge. *Biochem. J.* **255**, 285–290 (1988).
- Segal, A. W., Geisow, M., Garcia, R., Harper, A. & Miller, R. The respiratory burst of phagocytic cells is associated with a rise in vacuolar pH. *Nature* **290**, 406–409 (1981).
- Galvez, A. et al. Purification and characterization of a unique, potent, peptidyl probe for the high conductance calcium-activated potassium channel from venom of the scorpion *Buthus tamulus*. *J. Biol. Chem.* **265**, 11083–11090 (1990).
- Sanchez, M. & McManus, O. B. Paxilline inhibition of the alpha-subunit of the high-conductance calcium-activated potassium channel. *Neuropharmacology* **35**, 963–968 (1996).
- Clapp, L. H. & Tinker, A. Potassium channels in the vasculature. *Curr. Opin. Nephrol. Hypertens.* **7**, 91–98 (1998).
- Garcia, M. L. et al. Use of toxins to study potassium channels. *J. Bioenerg. Biomembr.* **23**, 615–646 (1991).
- Santos-Silva, J. E., Santos-Silva, M. C., Cunha, F. D. Q. & Assreuy, J. The role of ATP-sensitive potassium channels in neutrophil migration and plasma exudation. *J. Pharmac. Exp. Ther.* **300**, 946–951 (2002).
- Maingret, F., Patel, A. J., Lazdunski, M. & Honoré, E. The endocannabinoid anandamide is a direct and selective blocker of the background K⁺ channel TASK-1. *EMBO J.* **20**, 47–54 (2001).
- Lawson, K. Potassium channel openers as potential therapeutic weapons in ion channel disease. *Kidney Int.* **57**, 838–845 (2000).
- Fernandez-Fernandez, J. M., Nobles, M., Currid, A., Vazquez, E. & Valverde, M. A. Maxi K⁺ channel mediates regulatory volume decrease response in a human bronchial epithelial cell line. *Am. J. Physiol. Cell Physiol.* **283**, C1705–C1714 (2002).
- DeCoursey, T. E., Morgan, D. & Cherny, V. V. The voltage dependence of NADPH oxidase reveals why phagocytes need proton channels. *Nature* **422**, 531–534 (2003).
- Schrenzel, J. et al. Electron currents generated by the human phagocyte NADPH oxidase. *Nature* **392**, 734–737 (1998).
- Petrecchia, D. C., Nauseef, W. M. & Clark, R. A. Respiratory burst of normal human eosinophils. *J. Leukoc. Biol.* **41**, 283–288 (1987).
- Henderson, L. M., Chappell, J. B. & Jones, O. T. Internal pH changes associated with the activity of NADPH oxidase of human neutrophils. Further evidence for the presence of an H⁺ conducting channel. *Biochem. J.* **251**, 563–567 (1988).
- Cherny, V. V., Henderson, L. M., Xu, W., Thomas, L. L. & DeCoursey, T. E. Activation of NADPH oxidase-related proton and electron currents in human eosinophils by arachidonic acid. *J. Physiol. (Lond.)* **535**, 783–794 (2001).
- Kaczorowski, G. J. et al. High-conductance calcium activated potassium channel; structure, pharmacology, and function. *J. Bioenerg. Biomembr.* **28**, 255–267 (1996).
- Mahomed, G. A. & Anderson, R. Activation of human neutrophils with chemotactic peptide, opsonised zymosan and the calcium ionophore A23187, but not with a phorbol ester, is accompanied by efflux and store-operated influx of calcium. *Inflammation* **24**, 559–569 (2000).
- Jankowski, A. & Grinstein, S. A noninvasive fluorimetric procedure for measurement of membrane potential. Quantification of the NADPH oxidase-induced depolarization in activated neutrophils. *J. Biol. Chem.* **274**, 26098–26104 (1999).
- Tsien, R. Y. Fluorescent indicators of ion concentrations. *Methods Cell Biol.* **30**, 127–156 (1989).
- Vergun, O. et al. Exploration of the role of reactive oxygen species in glutamate neurotoxicity in rat hippocampal neurones in culture. *J. Physiol. (Lond.)* **531**, 147–163 (2001).
- Demaurex, N., Monod, A., Lew, D. P. & Krause, K. H. Characterization of receptor-mediated and store-regulated Ca²⁺ influx in human neutrophils. *Biochem. J.* **297**, 595–601 (1994).
- Grinstein, S. & Furuya, W. Cytoplasmic pH regulation in phorbol ester-activated human neutrophils. *Am. J. Physiol.* **251**, C55–C65 (1986).
- Willoughby, D., Thomas, R. & Schwenning, C. The effects of intracellular pH changes on resting cytosolic calcium in voltage-clamped snail neurones. *J. Physiol. (Lond.)* **530**, 405–416 (2001).
- Salvemini, D., Riley, D. P. & Cuzzocrea, S. SOD mimetics are coming of age. *Nature Rev. Drug Discov.* **1**, 367–374 (2002).
- Pick, E. & Keisari, Y. A simple colorimetric method for the measurement of hydrogen peroxide produced by cells in culture. *J. Immunol. Methods* **38**, 161–170 (1980).

Acknowledgements We thank P. Rich for helpful discussions, S. Ranasinghe for technical help with the FLIPR, and A. Scott for the illustrations. Financial support was provided by the Wellcome Trust and Medical Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.W.S. (t.segal@ucl.ac.uk).

OXI1 kinase is necessary for oxidative burst-mediated signalling in *Arabidopsis*

Maike C. Rentel^{1*}, David Lecourieux², Fatma Ouaked², Sarah L. Usher³, Lindsay Petersen^{1,4}, Haruko Okamoto¹, Heather Knight¹, Scott C. Peck⁵, Claire S. Grierson³, Heribert Hirt² & Marc R. Knight¹

¹Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

²Max F. Perutz Laboratories of the University of Vienna and Gregor-Mendel-Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna Biocenter, 1030 Vienna, Austria

³School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK

⁴Department of Molecular and Cell Biology, University of Cape Town, Private Bag Rondebosch 7701, South Africa

⁵Sainsbury Laboratory, Norwich NR4 7UH, UK

* Present address: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143-0450, USA

Active oxygen species (AOS) generated in response to stimuli and during development can function as signalling molecules in eukaryotes, leading to specific downstream responses^{1,2}. In plants these include such diverse processes as coping with stress (for example pathogen attack³, wounding⁴ and oxygen deprivation⁵), abscisic-acid-induced guard-cell closure⁶, and cellular development (for example root hair growth⁷). Despite the importance of signalling via AOS in eukaryotes, little is known about the protein components operating downstream of AOS that mediate any of these processes. Here we show that expression of an *Arabidopsis thaliana* gene (*OXI1*) encoding a serine/threonine kinase is induced in response to a wide range of H₂O₂-generating stimuli. OXI1 kinase activity is itself also induced by H₂O₂ *in vivo*. OXI1 is required for full activation of the mitogen-activated protein kinases (MAPKs) MPK3 and MPK6 after treatment with AOS or elicitor and is necessary for at least two very different AOS-mediated processes: basal resistance to *Peronospora parasitica* infection, and root hair growth. Thus, OXI1 is an essential part of the signal transduction pathway linking oxidative burst signals to diverse downstream responses.

Our attempts at identifying signalling components acting downstream of AOS signals focused on protein kinases as likely candidates. To identify the kinase genes specifically involved in AOS signalling in *Arabidopsis*, we compared kinase-encoding transcript levels in H₂O₂-treated (to mimic an oxidative burst) and control *Arabidopsis* seedlings by differential display of amplified complementary DNA fragments. In this screen, kinase-encoding cDNA sequences were amplified by polymerase chain reaction (PCR) with degenerate primers complementary to sequences conserved between protein kinases (data not shown). One gene identified in this way contained a coding region corresponding to a putative serine/threonine kinase (At3g25250) and was named *OXI1*, for *OXIDATIVE SIGNAL-INDUCIBLE1*.

OXI1 induction in response to AOS (H₂O₂) was confirmed by northern blot analysis (Fig. 1a). H₂O₂-treated seedlings were viable (data not shown), confirming previous reports that this stimulus was not lethal^{6,8}. As described above, AOS signalling in plants is triggered by a range of environmental stresses^{3–6}. We therefore examined the expression pattern of *OXI1* in response to a variety of stress treatments. *OXI1* was expressed to relatively high levels, particularly after wounding and treatment with cellulase (Fig. 1b). The prolonged kinetics of *OXI1* expression seen in response to H₂O₂ and cellulase contrasted with the rapid transient seen in response to wounding. These differences are probably due to the constant presence of the H₂O₂ and cellulase signals in comparison with the

discrete nature of the wounding signal, and the smaller number of cells affected in the latter case. Spatial patterns of transcription of the *OXI1* promoter in response to these two treatments were measured with an *OXI1::GUS* reporter gene fusion construct (Fig. 1c). GUS (β -glucuronidase) staining of the *OXI1::GUS* fusion lines after treatment with cellulase showed that the promoter was activated across the entire surface of the cotyledons (Fig. 1c, middle panel) that was consistent with the uniform degradation of cellulose by this enzyme. Conversely, wounding with a sharp razor blade increased GUS levels only along the cut edge of the wounded tissue and along the severed vascular tissues (Fig. 1c, right panel). These specific regions have been shown to produce H_2O_2 in response to wounding in plants⁴ including *Arabidopsis* (Supplementary Fig. 1; ref. 9). Control (untreated) seedlings did not show any staining (Fig. 1c, left panel) in cotyledons or leaves. We also found that *OXI1::GUS* fusion lines exhibited GUS expression along growing fungal hyphae of the virulent *P. parasitica* isolates Maks9 and Emco5, particularly in the adjacent layer of two to four cells and in cells penetrated by fungal haustoria (Fig. 2a). The staining pattern indicates that AOS production might be limited to cells in close proximity to the invading fungus. This is consistent with AOS production during pathogenesis^{10,11}, in which AOS generation in cells adjacent to virulent fungi has been shown¹².

In addition to stress-induced expression of *OXI1* in the aerial parts of the plant, the *OXI1* promoter is always active in the root. Here it is more highly expressed under growth conditions imposing a mild stress on the root, for example when seedlings are grown on ordinary agar instead of refined agar (Phytigel) (Fig. 3a; Supplementary Fig. 2), and often to a particularly high level in root hair cells. This would be consistent with the necessity of an AOS signal for normal root hair cell development⁷. Taken together, these data indicate that *OXI1* expression might be mediated by AOS produced in plant cells in response to stress and during root hair development, and they point to a role for the OXI1 protein in responses involving AOS.

To determine the specific function of OXI1 in AOS-mediated responses, we isolated an *oxi1*-null mutant line from the Wisconsin T-DNA insertion populations¹³ by using PCR-based screening (data not shown). The *oxi1* mutant expresses only a transcript encoding a truncated protein (the first 23 residues) lacking the catalytic site (data not shown), which is therefore non-functional. Complemen-

tation of the *oxi1* mutant with the wild-type *OXI1* gene confirmed that the T-DNA insertion created a null mutation and that this gene was responsible for all the phenotypes described below.

The necessity of OXI1 for resistance to a virulent pathogen was confirmed by comparing the susceptibility of the *oxi1* mutant and the wild type to pathogen infection. Assessment of the extent of fungal sporulation showed that *oxi1* seedlings have a significantly enhanced susceptibility to the virulent *P. parasitica* isolate Emco5 in comparison with the Ws-2 wild type (Fig. 2b and c). Resistance to the avirulent isolate Emoy2 was not removed in the *oxi1* mutant (data not shown). Because AOS are also likely to be produced in response to avirulent fungi, this implies that OXI1 either is not involved in host-specific resistance or is a redundant component in this context.

Cessation of root hair growth is a characteristic of some *rhd2* mutants¹⁴ that fail to generate necessary AOS during root hair development⁷. We therefore examined the *oxi1* mutant for altered root hair growth. The average root hair length of the *oxi1*-null mutant is significantly less than that of the Ws-2 wild type (226.2 and 300.3 μ m, respectively). There was also a greater proportion of very short root hairs in *oxi1* than in the wild type; this effect was more marked for roots growing through air than for those in contact with the agar (Fig. 3b and 3c; data not shown).

To test directly whether OXI1 kinase activity *in vivo* was induced by H_2O_2 , and hence to show that OXI1 acts downstream of AOS, OXI1 protein kinase activity was measured in *Arabidopsis* leaves (Fig. 4a). The addition of H_2O_2 resulted in strong activation of OXI1 at 5 min of treatment. Cellulase treatment also induced OXI1 activity in *Arabidopsis* leaves, although to a much higher level and over a longer period than in response to H_2O_2 . When leaves of *oxi1* mutant plants were analysed, OXI1 protein could not be detected, and activation of OXI1 by both H_2O_2 and cellulase was completely abrogated. These data show that OXI1 is a protein kinase that is activated in response to H_2O_2 and cellulase.

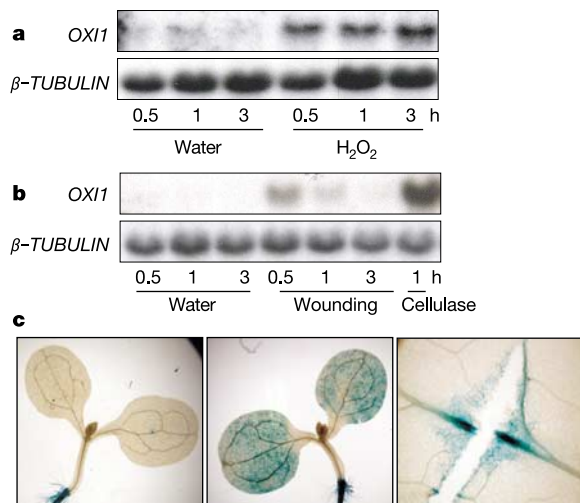


Figure 1 H_2O_2 , cellulase and wounding increase the expression of *OXI1*. **a**, Induction of *OXI1* in seedlings treated with H_2O_2 (10 mM) or water. **b**, Induction of *OXI1* by cellulase (0.01% w/v) or wounding. *OXI1* expression levels were determined by northern analysis. **c**, Seedlings containing an *OXI1::GUS* gene construct were incubated in water (left panel) or 0.1% w/v cellulase solution (middle panel). In the right panel, cotyledons were cut across the surface with a sharp razor blade. After 3 h, tissues were stained for GUS expression²⁵.

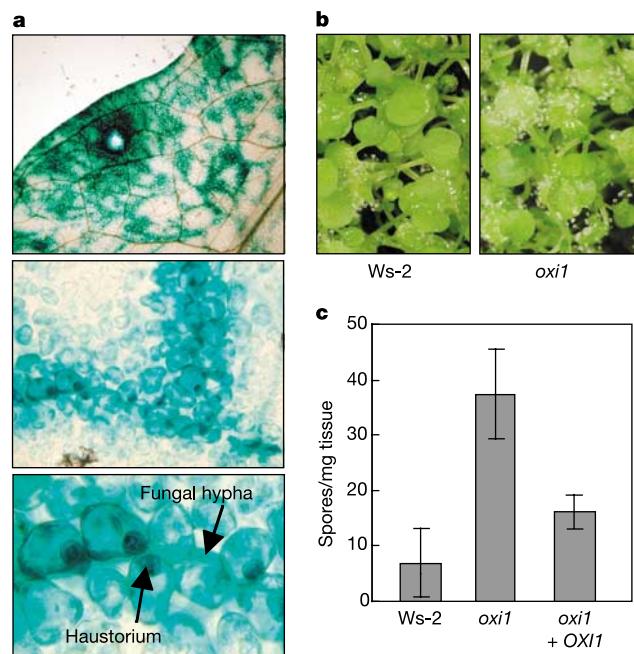


Figure 2 *OXI1* is necessary for basal resistance to *P. parasitica*. **a**, Plants containing an *OXI1::GUS* gene construct were infected with virulent *P. parasitica* (Maks9) and stained for GUS expression²⁵ 7 days after infection. Infected leaves show induction along fungal hyphae (top panel) and in cells containing fungal haustoria (middle panel, close-up in bottom panel). **b**, Infection levels of *oxi1* and Ws-2 wild-type seedlings 7 days after infection. **c**, Seedlings were infected with Emco5 and the level of fungal infection was assessed by determining the extent of sporulation on wild-type, *oxi1* mutant and *oxi1* complemented with wild-type *OXI1* (*oxi1* + *OXI1*) 6 days after infection (means $\pm$ s.e.m., $n = 3$).

Activation of two of the *Arabidopsis* MAPKs—MPK3 and MPK6—is known to be essential for signal transduction in response to H_2O_2 (ref. 15) and elicitor treatment¹⁶. MPK6 is also activated on wounding¹⁷, and SIMK, the alfalfa homologue of MPK6, regulates root hair growth¹⁸. To gain insight into the molecular targets of OXI1, H_2O_2 - and cellulase-triggered activation of these kinases was investigated in the *oxi1* mutant. In the wild type, both MPK3 and MPK6 displayed increased activity after treatment with H_2O_2 and cellulase (Fig. 4b and c, respectively). However, in the *oxi1* mutant H_2O_2 - and cellulase-triggered activation of both MPK3 and MPK6 was substantially reduced. Hence, OXI1 is needed for the full activation of MPK3 and MPK6 in response to AOS, and the phenotypes of the *oxi1* mutant might be attributable to reduced signalling through these MAPKs. In agreement with this notion, overexpression of an OXI1-YFP-c-Myc fusion protein in protoplasts increased H_2O_2 -induced MPK3 activity (data not shown), supporting the role for OXI1 in mediating the AOS-induced activation of the MAPKs. It is important to note that OXI1 gene expression and OXI1 and MPK3/6 kinase activities were upregulated more strongly in response to cellulase than to H_2O_2 . This might indicate that H_2O_2 treatment does not fully mimic a genuine oxidative burst generated by a stress signal such as cellulase and is therefore less effective. Timing, magnitude and spatial localization of the AOS signal in a real oxidative burst might be important in terms of activating downstream cellular processes. Alternatively, cellulase treatment might produce additional (to AOS) signals within the cell, for example by acting as an elicitor, and these other signals might also be capable of upregulating OXI1 gene expression and activity of the OXI1, MPK3 and MPK6 kinases.

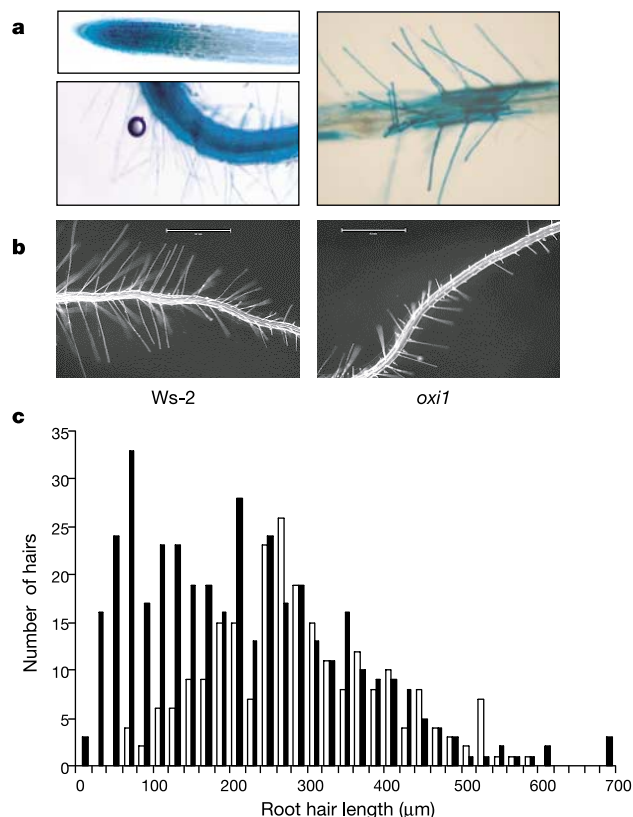


Figure 3 OXI1 is necessary for normal root hair development. **a**, Seedlings containing an OXI1::GUS gene construct were grown on 0.8% agar and subsequently stained for GUS expression. GUS was always expressed in the root, often to particularly high levels in root hair cells. **b**, Visual comparison of *oxi1* and Ws-2 wild-type roots grown through air. **c**, Distribution of mature root hair lengths of *oxi1* seedlings (filled bars) and Ws-2 wild-type seedlings (open bars) with roots growing through air. For complementation data, see Supplementary Fig. 3.

The findings described above indicate that the AOS-inducible OXI1 gene, encoding a serine/threonine kinase, is transcriptionally upregulated in response to wounding and pathogen attack, stimuli that are associated with H_2O_2 generation^{3,4}. In addition, the OXI1 promoter is active in roots, particularly in root hairs under growth conditions imposing a mild stress. H_2O_2 is uncharged and relatively stable, properties that allow this molecule to cross membranes and reach neighbouring cells, as shown in tobacco epidermal peels¹⁹. H_2O_2 might therefore function as a local signalling molecule. Accordingly, wounding and infection with *P. parasitica* increased expression of the OXI1::GUS reporter construct in several cell layers adjacent to the wounded or infected cells (Figs 1c and 2a, respectively). The OXI1 protein is essential for basal resistance against a virulent pathogen, because the *oxi1*-null mutant showed an enhanced susceptibility phenotype with an elevated production of sporangiophores (Fig. 2b, c). OXI1 is also necessary for normal root hair development during stress (Fig. 3b, c), root hair development itself requiring AOS generated by the *Arabidopsis* NADPH oxidase AtrbohC/RHD2 (ref. 7). Notably, OXI1 falls into the same kinase gene family as the INCOMPLETE ROOT HAIR ELONGATION (IRE) gene, which is also involved in root hair growth²⁰. The OXI1 protein has protein kinase activity that is induced *in vivo* by H_2O_2

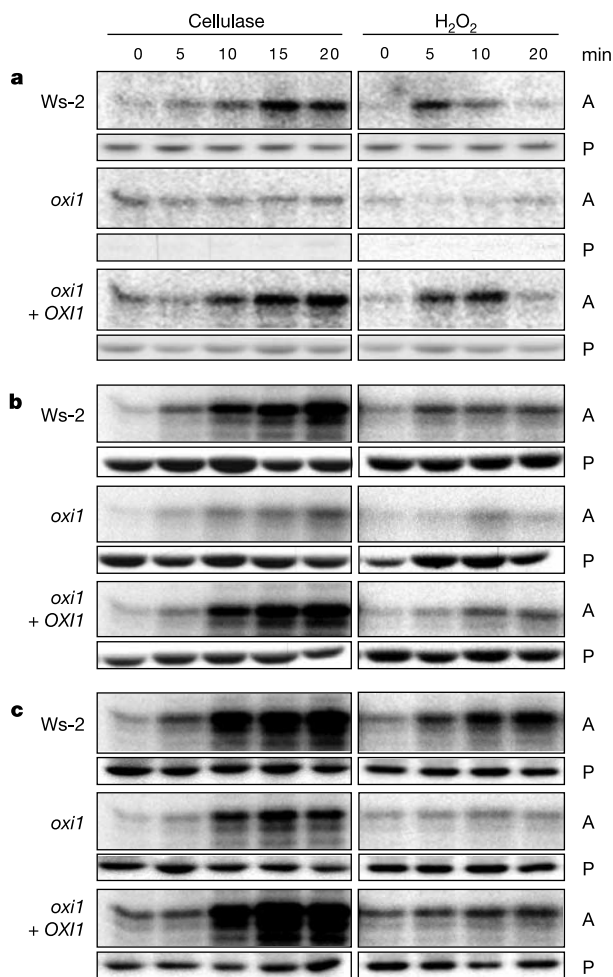


Figure 4 OXI1 kinase activity is induced by H_2O_2 and cellulase, and OXI1 is necessary for MPK3 and MPK6 activation. **a**, Cellulase and H_2O_2 induce activity of the native OXI1 kinase, immunoprecipitated with an anti-OXI1 antibody from *Arabidopsis* leaves. **b**, **c**, MPK3 (**b**) and MPK6 (**c**) activity in cell extracts, assessed by immunocomplex kinase assays. Samples were treated with H_2O_2 (10 mM) or cellulase (0.1% w/v). Myelin basic protein was used as a substrate. A, activity; P, protein levels, determined by western blotting with antibodies specific for OXI1 (**a**), MPK3 (**b**) and MPK6 (**c**).

and cellulase (Fig. 4a). Downstream (indirect) targets of OXI1 are likely to include MPK3 and MPK6, because *OXI1* is required for full activation of these kinases (Fig. 4b, c). *OXI1* gene expression is also induced by cold, by osmotic stress and by heat (data not shown). It is therefore possible that OXI1 is involved in many more AOS-dependent processes and might act as a universal mediator of oxidative bursts and stresses in *Arabidopsis*. □

Methods

Plant growth conditions

Arabidopsis thaliana plants were grown on 1 × MS medium and 0.8% w/v agar or in peat plugs at 21 °C under 16-h light conditions at 60 μmol m⁻² s⁻¹ light intensity. For infection with *P. parasitica*, plants on peat plugs were placed under a transparent dome at 16 °C under 8-h light conditions.

OXI1::GUS reporter gene and OXI1 complementation constructs

A 1,610-base-pair (bp) fragment containing the promoter and the first two nucleotides of the *OXI1* coding sequence was amplified from genomic DNA by PCR with the primers 5'-GCCGCGATCCCGCTGGGATAATCTCAAAGG-3' and 5'-CGCGCTGCAGATAA TGTCGACGTTAGTAAAC-3', and cloned into the pDH51 plasmid²¹ containing the *GUS* reporter gene. The *OXI1::GUS* fragment including a cauliflower mosaic virus terminator sequence was subsequently excised by restriction with *Bam*HI and *Kpn*I and ligated into the *Agrobacterium* binary vector pBIN19 (ref. 22). A 4,704-bp fragment containing the wild-type *OXI1* gene as well as the 5' 2,148 bp and the 3' 1,173 bp was amplified from genomic DNA by PCR with the primers 5'-GCGCGAATTCTAAAGTGTAGGCGAATA GCTGGAGACT-3' and 5'-GCGCGGATCCGCTATCATTATTAGGAGAATGGGAG ATTG-3'. The fragment was cloned into the pFGC5941 binary plasmid by restriction with *Eco*RI and *Bam*HI and subsequent ligation. The resulting plasmids were transformed into *Agrobacterium tumefaciens*, strain C58C1 (ref. 23). Col-0, Ws-2 (wild types) and *oxi1* mutant plants were transformed by the dipping method²⁴.

GUS staining

Seven-day-old seedlings were stained for GUS activity by submersion in staining solution (modified from ref. 25: 50 mM NaPO₄ pH 7.2, 0.5 mM K₃Fe(CN)₆, 0.5 mM K₄Fe(CN)₆, 10 mM EDTA, 0.01% (v/v) Triton X-100, 2 mM 5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid substrate in dimethylformamide (diluted from a 100 mM stock solution) and vacuum infiltration to 50 kPa for 3 min. Samples were incubated at 37 °C until the desired staining intensity was reached, and were destained in 80% ethanol. Images were taken with a Nikon Coolpix 990 digital camera mounted on a Leica DM R microscope.

Northern blot analysis

Northern analysis was performed as described⁸. Seven-day-old *Arabidopsis* seedlings were transferred into water and left to recover for 2.5 h. H₂O₂ or cellulase (Onozuka R-10; Yakult Ltd, Tokyo, Japan) was added to the indicated final concentrations. For wounding, about 25% of the cotyledon surface was damaged with tweezers. Control samples remained in water. Samples were frozen after 0.5, 1 or 3 h. *OXI1* and β-TUBULIN (*TUB*)-specific probes were labelled with ³²P and hybridized to total RNA blotted onto nylon membrane.

Assessment of *P. parasitica* sporulation

Seven-day-old seedlings or 4-week-old plants on peat plugs were sprayed with a spore suspension of 5 × 10⁴ spores ml⁻¹. At 6 days after infection, 150–300 mg tissue (cotyledons and small leaves) from separate plugs was removed and weighed, and 200 μl H₂O was added. After vortex-mixing for 30 s, spores were counted on a haemocytometer. The mean of ten squares (each 0.1 mm² suspension) was recorded per sample and samples were scored twice to ensure accuracy. The values were then converted to the number of spores per mg fresh weight²⁶.

Imaging and measurement of root hairs

Images of roots growing in air, away from the surface of solid growth medium (0.8% Phytigel; adapted from ref. 27) were captured with a Leica MZFLIII stereo binocular microscope, a SPOT RT colour digital camera, and SPOT Advanced software version 3.1 (Diagnostic Instruments). Root hairs that were in focus throughout their length were measured with Image J 1.29x (W. Rasband) and the results were analysed with Microsoft Excel 2000 and Minitab.

Immunocomplex kinase assays

Kinase activity assays were performed as described previously²⁸ on extracts from 3-week-old leaves. In brief, cleared cell extracts containing equal protein amounts were subjected to a 2 h preincubation in the presence of 20 μl mixed protein A–Sepharose and protein G–Sepharose beads (1:1). The supernatant was immunoprecipitated with 5 μg protein A-purified antibodies raised against synthetic peptides encoding the carboxy-terminal 7 or amino-terminal 15 amino acids of MPK6 and MPK3, respectively, and 20 μl of protein A–Sepharose beads. The antibody to OXI1 kinase was raised against a synthetic peptide encoding the C-terminal 15 amino acids. The beads were washed three times with wash buffer and once with kinase buffer. Kinase reactions of the immunoprecipitated proteins were performed at 24 °C for 30 min in 15 μl kinase buffer containing 5 μg myelin basic protein, 0.1 mM ATP and 2 μCi [γ-³²P]ATP. The reactions were stopped by adding 4 × SDS loading buffer. Phosphorylation of myelin basic protein as substrate for MAPKs was analysed by autoradiography after SDS–polyacrylamide-gel electrophoresis.

Western blot analysis

Western blotting was performed with equal amounts (20 μg) of protein extracts separated by SDS–polyacrylamide-gel electrophoresis, immunoblotted to poly(vinylidene difluoride) membranes (Millipore) and probed with affinity-purified antibodies either in Tris-buffered saline pH 7.4 containing 0.05% Tween 20 for MAPKs or in phosphate-buffered saline pH 7.4 containing Tween 20 for OXI1. Alkaline-phosphatase-conjugated goat anti-rabbit IgG (Sigma) was used as secondary antibody and the reaction was revealed by fluorography (CDP-Star; Amersham Life Sciences).

Received 16 August 2003; accepted 19 January 2004; doi:10.1038/nature02353.

- Finkel, T. Oxygen radicals and signaling. *Curr. Opin. Cell Biol.* **10**, 248–253 (1998).
- Hancock, J. T., Desikan, R. & Neill, S. J. Role of reactive oxygen species in cell signalling pathways. *Biochem. Soc. Trans.* **29**, 345–350 (2001).
- Levine, A., Tenhaken, R., Dixon, R. & Lamb, C. H₂O₂ from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell* **79**, 583–593 (1994).
- Orozco-Cárdenas, M. L., Narvaez-Vasquez, J. & Ryan, C. A. Hydrogen peroxide acts as a second messenger for the induction of defense genes in tomato plants in response to wounding, systemin, and methyl jasmonate. *Plant Cell* **13**, 179–191 (2001).
- Baxter-Burrell, A., Chang, R., Springer, P. & Bailey-Serres, J. Gene and enhancer trap transposable elements reveal oxygen deprivation-regulated genes and their complex patterns of expression in *Arabidopsis*. *Ann. Bot.* **91** (special number), 129–141 (2003).
- Pei, Z.-M. et al. Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells. *Nature* **406**, 731–734 (2000).
- Foreman, J. et al. Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* **422**, 442–446 (2003).
- Knight, H., Brandt, S. & Knight, M. R. A history of stress alters drought calcium signalling pathways in *Arabidopsis*. *Plant J.* **16**, 681–687 (1998).
- Orozco-Cárdenas, M. & Ryan, C. A. Hydrogen peroxide is generated systemically in plant leaves by wounding and systemin via the octadecanoid pathway. *Proc. Natl Acad. Sci. USA* **96**, 6553–6557 (1999).
- Hückelhoven, R., Fodor, J., Preis, C. & Kogel, K.-H. Hypersensitive cell death and papilla formation in barley attacked by the powdery mildew fungus are associated with hydrogen peroxide but not with salicylic acid accumulation. *Plant Physiol.* **119**, 1251–1260 (1999).
- Torres, M. A., Dangl, J. L. & Jones, J. D. *Arabidopsis* gp91phox homologues *AtbbohD* and *AtbbohF* are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc. Natl Acad. Sci. USA* **99**, 517–522 (2002).
- Borden, S. & Higgins, V. J. Hydrogen peroxide plays a critical role in the defence response of tomato to *Cladosporium fulvum*. *Physiol. Mol. Plant Pathol.* **61**, 227–236 (2002).
- Krysan, P. J., Young, J. C. & Sussmann, M. R. T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* **11**, 2283–2290 (1999).
- Parker, J. S., Cavell, A. C., Dolan, L., Roberts, K. & Grierson, C. S. Genetic interactions during root hair morphogenesis in *Arabidopsis*. *Plant Cell* **12**, 1961–1974 (2000).
- Kovtun, Y., Chiu, W.-L., Tena, G. & Sheen, J. Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl Acad. Sci. USA* **97**, 2940–2945 (2000).
- Asai, T. et al. MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* **415**, 977–983 (2002).
- Ichimura, K., Mizoguchi, T., Yoshida, R., Yuasa, T. & Shinozaki, K. Various abiotic stresses rapidly activate *Arabidopsis* MAP kinases *AtMPK4* and *AtMPK6*. *Plant J.* **24**, 655–665 (2000).
- Samaj, J. et al. Involvement of the mitogen-activated protein kinase SIMK in regulation of root hair tip growth. *EMBO J.* **21**, 3296–3306 (2002).
- Allan, A. C. & Fluhr, R. Two distinct sources of elicited reactive oxygen species in tobacco epidermal cells. *Plant Cell* **9**, 1559–1572 (1997).
- Oyama, T., Shimura, Y. & Okada, K. The *IRE* gene encodes a protein kinase homologue and modulates root hair growth in *Arabidopsis*. *Plant J.* **30**, 289–299 (2002).
- Pietrzak, M., Shillito, R. D., Hohn, T. & Potrykus, I. Expression in plants of two bacterial antibiotic-resistance genes after protoplast transformation with a new plant expression vector. *Nucleic Acids Res.* **14**, 5857–5868 (1986).
- Bevan, M. W. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* **12**, 8711–8721 (1984).
- Deblacere, R. et al. Efficient octopine T₁ plasmid-derived vectors for *Agrobacterium*-mediated gene-transfer to plants. *Nucleic Acids Res.* **13**, 4777–4788 (1985).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
- Martin, T., Schmidt, R., Altmann, T. & Frommer, W. Non-destructive assay systems for detection of β-glucuronidase activity in higher plants. *Plant Mol. Biol. Reprt* **10**, 37–46 (1992).
- Weigel, D. & Glazebrook, J. *Arabidopsis – a Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).
- Wymier, C. L., Bibikova, T. N. & Gilroy, S. Cytoplasmic free calcium distributions during the development of root hairs of *Arabidopsis thaliana*. *Plant J.* **12**, 427–439 (1997).
- Cardinale, F., Meskine, I., Ouaked, F. & Hirt, H. Protein kinase signaling pathways at the level of two distinct mitogen-activated protein kinase kinases. *Plant Cell* **14**, 703–711 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Rehmany for the *P. parasitica* isolates; A. Hailey for statistical analysis; B. Ellis and G. Miles for discussions; and N. Evans, S. Gurr and R. Capper and K. Denby for practical and moral support. This work was funded by the BBSRC, the Gatsby Foundation, the EU TMR program, the Cannon Collins Educational Trust for Southern Africa (CCETSA), the South African National Research Foundation (NRF), and the Austrian Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.R. (mrentel@cmp.ucsf.edu) or H.H. (heribert.hirt@univie.ac.at).

Visualization of release factor 3 on the ribosome during termination of protein synthesis

Bruno P. Klaholz^{1,2}, Alexander G. Myasnikov³ & Marin van Heel¹

¹Department of Biological Sciences, Imperial College London, London SW7 2AY, UK

²Department of Structural Biology and Genomics, Institute of Genetics and Molecular and Cellular Biology, CNRS/INSERM/ULP, BP 10142, 67404 Illkirch, France

³Department of Molecular Biology, Belozersky Institute, Building A, Moscow State University, Moscow 119899, Russia

Termination of protein synthesis by the ribosome requires two release factor (RF) classes. The class II RF3 is a GTPase that removes class I RFs (RF1 or RF2) from the ribosome after release of the nascent polypeptide^{1–3}. RF3 in the GDP state binds to the ribosomal class I RF complex, followed by an exchange of GDP for GTP and release of the class I RF. As GTP hydrolysis triggers release of RF3 (ref. 4), we trapped RF3 on *Escherichia coli* ribosomes using a nonhydrolysable GTP analogue. Here we show by cryo-electron microscopy that the complex can adopt two different conformational states. In ‘state 1’, RF3 is pre-bound to the ribosome, whereas in ‘state 2’ RF3 contacts the ribosome GTPase centre. The transfer RNA molecule translocates from the peptidyl site in state 1 to the exit site in state 2. This translocation is associated with a large conformational rearrangement of the ribosome. Because state 1 seems able to accommodate

simultaneously both RF3 and RF2, whose position is known from previous studies^{5,6}, we can infer the release mechanism of class I RFs.

Class I and class II RFs mediate the transition from a state stalled by the stop codon to a state that allows ribosome dissociation or sliding on messenger RNA and recycling^{1–3}. A combination of crystallography and cryo-electron microscopy (cryo-EM) has shown that class I RFs transmit the stop codon signal directly to the peptidyl transferase centre, where hydrolysis of the peptidyl (P) site transfer RNA–peptidyl bond is catalysed^{5,6}. By contrast, the location of the class II RF3 on the ribosome has remained unknown, and no crosslinking or footprinting data are available.

We have determined the structure of an RF3 ribosomal complex by single-particle cryo-EM and three-dimensional reconstruction. Incubation of *E. coli* ribosomes containing a deacetylated P site tRNA and a stop-codon mRNA, RF3 and the non-hydrolysable GTP analogue GTPNP enabled us to stabilize an intermediary complex before the dissociation of RF3 (ref. 4). Initial maps showed a clear mass of density for RF3; however, disordered regions in the maps, in particular in the beak region of the 30S subunit, indicated that there was structural heterogeneity in the sample.

Separation of different types of particles by image processing revealed two different functional states of the RF3-bound 70S ribosome, which were present in comparable amounts in the sample. The two structures differ by the position and conformation of RF3, by the position of the tRNA molecule, and by the conformational state of the whole ribosome. We found a direct correlation among conformational changes in distinct parts of the ribosome complex, such as the L1 stalk, 30S beak, 30S toe, tRNA and RF3 (see Methods and Supplementary Information). The interpretation of the maps is largely based on the available *Thermus thermophilus*

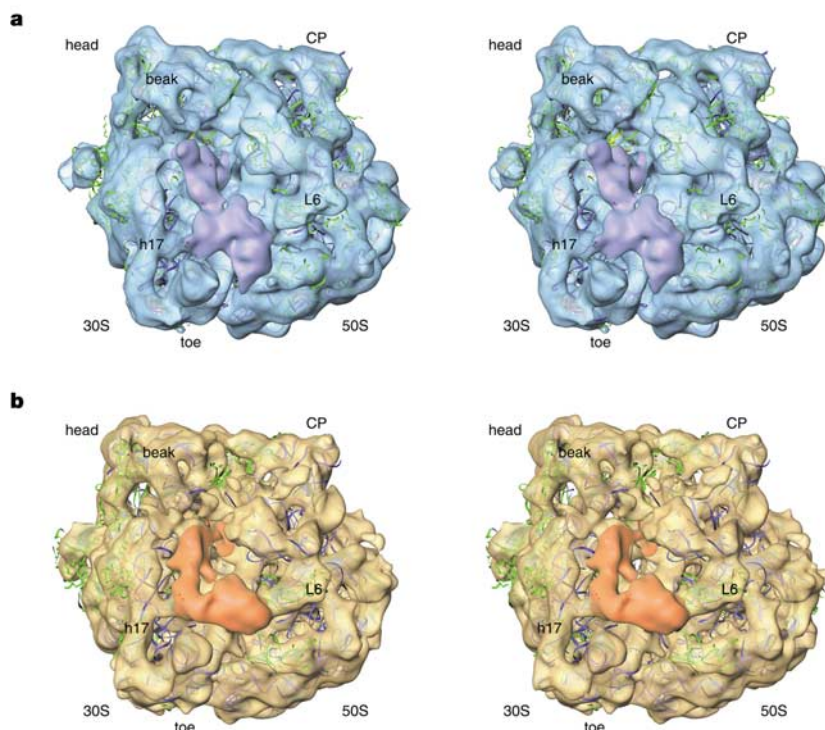


Figure 1 Stereo views of RF3–ribosome complexes seen from the ‘front’. Shown are cryo-EM density maps of the two 70S–RF3–GTPNP functional states separated by image processing: **a**, state 1 (light blue); **b**, state 2 (ochre). The cryo-EM density of RF3 is depicted in magenta (state 1) or orange (state 2). The dynamically fitted 70S crystal structure from *T. thermophilus*⁷ is shown inside the maps (purple, ribosomal RNA; green, ribosomal proteins). The subunits are labelled, and some ribosomal landmarks are indicated (the 30S head, beak, toe and helix h17, and the 50S central protuberance ‘CP’

and protein L6). In state 1, the conformation of 30S is similar to that in pre-translocation states found in previous studies; the RF3 G/G’ domain is oriented towards the toe of the 30S subunit, far from the GTPase centre on the 50S subunit. In state 2, the 30S subunit is rotated away from the observer (by ~6° compared with **a**); the RF3 G/G’ domain is seen in a position close to the GTPase centre, comparable to, for example, the position of the corresponding domains in EF-Tu and EF-G, as determined in previous cryo-EM studies.

crystal structures of the 70S ribosome⁷, and of the homologous parts of the elongation factor EF-G His573Ala mutant, in which domain III is well ordered⁸. The comparison of the two states suggests that 'state 1' represents a pre-translocation complex with a P-site tRNA and a pre-bound molecule of RF3 in an inactive conformation, whereas 'state 2' represents a post-translocation state in which RF3 is tightly bound to the GTPase centre and the tRNA has moved from the P to the exit (E) site (Figs 1–3).

The common feature of the two 70S–RF3–GDPNP complexes is the global binding site of RF3, which is similar to that of elongation factors EF-Tu^{9–11} and EF-G^{12,13}. Comparison of the complexes with the control 70S ribosome⁵ reveals, in both cases, a large density attributable to RF3 in the vicinity of the GTPase centre and the protein L6 area (Fig. 1). The density extends up to the 30S subunit, where it shows a kink towards the ribosomal intersubunit space, further extending into the aminoacyl (A) site and towards protein S12 (Figs 1 and 2). In both states the RF3 density is connected throughout, allowing individual domains, from the amino-terminal G domain to the carboxy terminus, to be assigned.

Although a perfect fit of the EF-G crystal structure^{8,14,15} is not expected because of sequence differences between RF3 and EF-G, the fitting helps to identify the individual domains that interact with the ribosome (Fig. 2a, b). The structural unit formed by domains I and II of EF-G, which contain the GTPase activity, can be assigned unambiguously; positioning domain III requires its rotation towards the A site with respect to domains I and II. The remaining density clearly corresponds to the C-terminal part of RF3, which has sequence similarity with the second part of EF-G domain IV (residues Pro 535 to Val 608) as evident from multiple sequence

alignments (see Methods and ref. 16). The conserved part of domain IV would correspond to a helix–turn–helix motif in EF-G, which has the correct size to fill the remaining cryo-EM density. As compared with EF-G, this domain would have to be rotated to maintain the connectivity to domain III. The first part of domain IV and the entire domain V of EF-G are missing in RF3 sequences.

The conformations of RF3 and the whole ribosome differ fundamentally in the two distinct states (Fig. 2c). In state 1, RF3 has only a few contact points with the ribosome including the 30S toe and shoulder, the decoding site and a weak contact to the tip of helix H44 of the 23S RNA (Figs 1a and 2a). Notably, the G domain does not contact the α -sarcin–ricin loop (SRL), which is the main component of the GTPase centre on the 50S subunit. Rather, it is found rotated away from the 50S subunit, with domains I and II located at the entrance of the intersubunit space. The only contact of domains I and II with the ribosome is that with the 30S toe, which could function as an intermediate anchor point for RF3. Domain III contacts the 30S subunit at the level of helices h16 and h17. Inside the intersubunit space, the C-terminal part of RF3 makes a strong contact with protein S12, and its density extends over helix h44 and reaches the tip of helix H69.

By contrast, RF3 in state 2 (Figs 1b and 2b) is tightly bound to the ribosome and has its G domain oriented towards the GTPase centre. The fit of EF-G domains I and II into the map places the guanosine nucleotide pocket directly juxtaposed to the SRL. Contacts to the C-terminal part of protein L6 are mediated completely through the G' domain, which is conserved between EF-G and RF3 (ref. 17) and provides an anchor point on the 50S subunit. Domain II interacts with the middle part of helix h17, whereas domain III contacts the area around helices h16 and h17 (16S RNA). There are several contacts to the area around the A site: domain III interacts with the SRL on the 50S subunit and with protein S12 on the 30S subunit. The tip of helix h5 (16S RNA) contacts RF3 between domains II and III. Notably, a strong density corresponding to the RF3 C-terminal domain reaches the tip of helix H69 and interconnects helices H69, H89 and H38 (A-site finger).

Two types of conformational changes in the 30S subunit occur on transition from state 1 to 2. First, there is a global rotation of the 30S subunit by roughly 6° with respect to the 50S subunit, with the rotation axis close to the area defined by helices H67 and H71 of the 23S RNA (intersubunit bridges B2 and B3, respectively). Most of the subunit interface contacts outside the H69 area are maintained, although they are shifted slightly. Second, there is a rearrangement of several areas, including the beak, toe and L1 areas (Figs 1 and 3). In state 2, the beak is folded against the 30S head and interacts strongly with protein S3, resulting in an increase in the gap between the head and shoulder of the 30S subunit (Fig. 1). In state 1, the L1 area is oriented towards the outside of the ribosome (Fig. 3), whereas in state 2 it moves inwards, thus stabilizing the E-site tRNA.

The conformational changes in the ribosome between states 1 and 2 that are induced by RF3 resemble those between the pre- and post-translocational states that are induced by EF-G^{13,18}. However, in contrast to the translocation induced by EF-G, where the A- and P-site tRNAs move to the P and E sites, respectively, RF3 removes the A-site-bound class I RF2 and induces translocation of tRNA from the P site to the E site (Fig. 3). The conformational change in RF3 from an open to a closed conformation, initiated at the level of the G domain (Fig. 2c), seems to be transmitted through domains III and IV to the tip of H69, which moves slightly into the P site owing to the absence of the tRNA (Fig. 3). Being close to the rotation centre of the 30S subunit, H69 could be a key for the structural transition of the ribosome and tRNA translocation. The fact that, under our experimental conditions, the tRNA can be translocated by RF3 would also explain how RF3 can substitute for EF-G in ribosomal recycling factor (RRF)-dependent recycling². The state-2 complex might be an appropriate substrate for RRF, according to its localization using footprinting¹⁹.

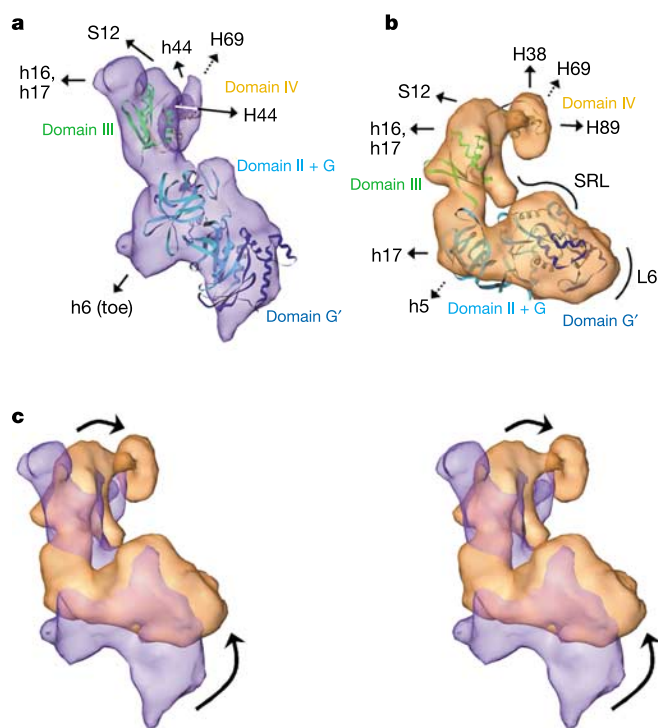


Figure 2 Interaction patterns of RF3 in the ribosome. The isolated RF3 densities, in their state-1 (magenta, **a**) and state-2 (orange, **b**) conformations, are viewed from the same overall direction as in Fig. 1. The manually fitted RF3-homologous parts of the *T. thermophilus* EF-G structure⁸ are shown (domains I, II and III, and the second part of domain IV, coloured as follows: dark blue, subdomain G'; light blue, subdomain G plus domain II; green, domain III; and yellow, domain IV). The 30S RNA helices are denoted by 'h', whereas the 50S helices are denoted by 'H'. **c**, Stereo view of the superposition of the state-1 and the state-2 RF3 densities, illustrating the three-dimensional structural transition of RF3 from an open to a closed conformation.

The position of the class I RF2 in a post-termination complex, the substrate for RF3, is known from our previous cryo-EM study⁵ and can be superimposed with the two RF3 complexes. In state 1, RF3 leaves a binding pocket completely free for RF2 domains 1 and 3, and for the largest part of domains 2 and 4 (Fig. 4). Steric clash with a small part of RF2 domains 2 and 4 is seen only where RF3 domain

IV crosses helix h44 of the 16S RNA. By contrast, in state 2 RF3 competes for the RF2-binding site, because RF3 occupies the area of RF2 domains 2 and 4, and parts of domains 1 and 3.

In state 1, the shape and position of the class I and class II RFs complement each other remarkably well, suggesting that there is a direct interaction within termination intermediates: in this model,

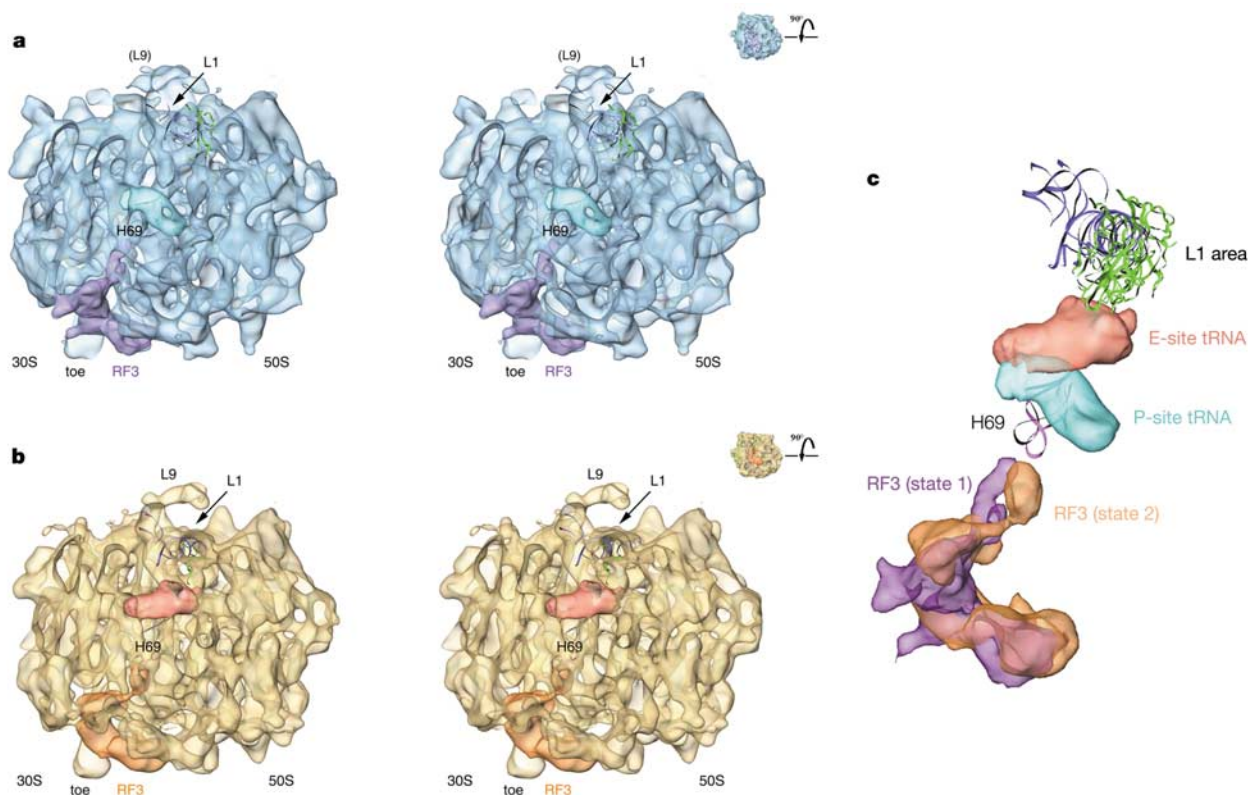


Figure 3 tRNA translocation and 70S conformational changes. **a, b**, Stereo views of the 70S–RF3 complexes in state-1 and state-2 conformations seen in a top-view orientation, with the upper part of the 70S removed for clarity (view rotated by 90° with respect to Fig. 1). The maps are coloured as in Figs 1 and 2, with the tRNA densities highlighted in cyan and red for states 1 and 2, respectively, the L1 protein structure in green, and the L1 RNA stalk in purple. The position of helix H69 is indicated; RF3 contacts the tip of H69 in both states. Note that in both maps protein L9 is positioned in an area close to where it was first found in the 50S structure²⁸, and not where it is found in the

T. thermophilus structure⁷, where it protrudes from the 50S subunit and is involved in a crystal contact. In the state-1 conformation, the (head of the) 30S subunit and the L1 stalk are rotated away from each other. By contrast, in the state-2 conformation, the 30S subunit and the L1 stalk are rotated towards each other. Note the proximity of the E-site tRNA (shown in red) and the L1 stalk and protein in this ‘post-translocational’ arrangement. **c**, Superposition of the essential parts of the stereo maps shown in **a** and **b** in the same viewing direction.

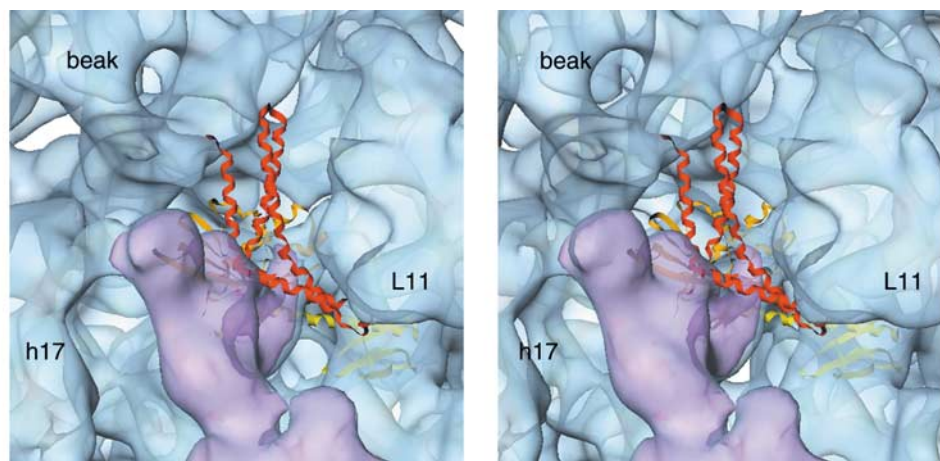


Figure 4 Superposition of RF2 onto the state-1 RF3 ribosomal complex. In this stereo pair, the spatial complementarity between RF3 in the state-1 complex (Fig. 1a) is illustrated by superposition with the model structure⁵ of RF2 (domain 1 of RF2 is shown in red, domains

2 and 4 in orange, and domain 3 in yellow). Virtually no steric hindrance exists, even without assuming structural changes owing to interactions between these RFs. By contrast, the RF3 position in state 2 is incompatible with the simultaneous presence of RF2.

the C-terminal part of domain III would be close to domain 1 of RF2, and domain IV of RF3 would interact with domains 2 and 4 of RF2 (Fig. 4). This model is in agreement with biochemical evidence showing that RF2 domain 1 is required for recycling by RF3 (ref. 20). Notably, the interaction of the C-terminal domains of eukaryotic class I and II RFs²¹ seems similar, considering that the C-terminal part of eRF1 is equivalent to the N-terminal part of RF2 (ref. 5). Dissociation of RF1 or RF2 could be induced by the transition from state 1 to state 2, which implies a hinge movement of domains III and IV with respect to domains I and II (Fig. 2c). The unique role of domain IV is in agreement with the RF3-specific activity of the C terminus, as suggested from multiple sequence alignments.

The presence of two structures within a sample of homogeneous composition actually raises the question of whether GDPNP is a true analogue of GTP: state 2 represents the functional complex after class I RF release, because RF3 overlaps with the class I RF binding site. Because GDPNP cannot be hydrolysed, however, RF3 remains bound on the ribosome. By contrast, the state-1 situation most probably represents the complex before class I RF release and therefore will not have GTP bound^{3,4,21}. State 1 is thus a state where either no nucleotide is bound (in which case it represents the intermediate during GDP/GTP exchange) or a GDPNP molecule behaving like a GDP analogue is bound. The high GDPNP concentration in the sample (see Methods) favours the latter explanation. □

Methods

Sample preparation

Sample and specimen preparations were done largely as described^{5,21}. *E. coli* ribosomes programmed with *MFTI* mRNA, which are paused with a UAA stop codon in the A site and a tetrapeptidyl-tRNA in the P site (called release complexes; RCs), were prepared by *in vitro* translation and isolated by gel filtration²¹. The occupancy of RF3 was determined by nitrocellulose filter binding of RF3·³H]GDPNP·RC to be ~80% (ref. 21). We incubated 5.2 pmol of RCs (in 50 µl) with 100 pmol RF3, 0.2 mM GDPNP (guanosine-5'-β,γ-imidotriphosphate) and 0.2 mM puromycin (to remove the tetrapeptide) for 10 min at 37 °C. The sample was stored on ice until use.

Cryo-EM and image processing

Samples were diluted with polymix buffer containing 2 µM RF3 and 0.2 mM GDPNP to a final concentration of 10 nM before application to a holey carbon grid. Excess solution was blotted and the grid was flash frozen in liquid ethane, resulting in ribosome complexes embedded in a thin film of vitrified buffer spanning the holes on the carbon film. RC alone was used as a reference². Low-dose micrographs were recorded with a Philips CM200 FEG cryo-microscope. Images were collected at X50,000 magnification and a defocus range of 1.0–2.8 µm.

Good micrographs (verified by optical diffraction) were digitized on an Image Science patchwork densitometer²² with a step size of 3 µm. The digitized images were coarsened by a factor of 4, which resulted in a pixel size corresponding to 2.4 Å at the specimen level. Single-molecule images (21,082) were selected interactively from the micrographs, and phase-contrast-transfer function correction was done by phase flipping on the raw data. All image processing was done in the context of the IMAGIC-V software system^{22,23}.

A variation was the processing, which converged towards two three-dimensional structures simultaneously to account for the heterogeneous population of particle images, enabling the analysis of two specific intermediate states (see Supplementary Information). The resolution of the final three-dimensional structures was determined with Fourier shell correlation to 17.9 Å, 13.3 Å, 23.8 Å and 17.5 Å, 14.4 Å, 24.7 Å, for states 1 and 2, respectively, according to the 0.14 (ref. 24), 3σ, 0.5 criteria. The high-frequency components of the final three-dimensional maps were suppressed by low-pass filtering at 15 Å.

Map fitting and interpretation

Detailed interpretation of the cryo-EM map of the RF3 *E. coli* ribosome complexes was done after dynamically fitting the global *T. thermophilus* ribosome crystal structure⁷ (Protein Data Bank (PDB) accession codes 1GIX and 1GIY) into the cryo-EM maps. This fitting was done in three steps: first, manual fitting of the global 70S ribosome as a rigid body using the program O (ref. 25); second, rigid-body fitting of the global 70S ribosome using the program CNS²⁶; and third, reciprocal-space torsion angle molecular dynamics and simulated annealing (start and end temperatures 1,500 K and 300 K, respectively; resolution range 200–18 Å, consistent with the resolution assessment according to the 0.14 criterion²⁴), implemented in CNS²⁶. For the last procedure, structural domains of the RNA were defined and maintained as interconnected rigid bodies, which considerably reduces the parameters to be refined because the interdomain hinges are the only torsion angles to be refined. The position of ribosomal proteins was harmonically restrained. We checked the resulting fit by visual inspection using the program O (ref. 25), and refitted some minor parts manually.

Multiple sequence alignments of 500 RF3-related sequences were done with the program PipeAlign (http://igbmc.u-strasbg.fr/PipeAlign; ref. 27). Pair-wise sequence comparison between RF3 (*E. coli*) and EF-G (*T. thermophilus*) yields 18% identity and 38% similarity overall; 27% and 51% for domain I; 15% and 36% for the G' domain; 20%

and 41% for domain II; 25% and 54% for domain III; and 9% and 34% for domain IV. Figures were prepared using the program Amira (TGS).

Received 9 October 2003; accepted 8 January 2004; doi:10.1038/nature02332.

- Freistoffer, D. V., Pavlov, M. Y., MacDougall, J., Buckingham, R. H. & Ehrenberg, M. Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J.* **16**, 4126–4133 (1997).
- Grentzmann, G. *et al.* Release factor RF-3 GTPase activity acts in disassembly of the ribosome termination complex. *RNA* **4**, 973–983 (1998).
- Kisselev, L., Ehrenberg, M. & Frolova, L. Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J.* **22**, 175–182 (2003).
- Zavialov, A. V., Buckingham, R. H. & Ehrenberg, M. A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell* **107**, 115–124 (2001).
- Klaholz, B. P. *et al.* Structure of the *Escherichia coli* ribosomal termination complex with release factor 2. *Nature* **421**, 90–94 (2003).
- Rawat, U. B. *et al.* A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature* **421**, 87–90 (2003).
- Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
- Laurberg, M. *et al.* Structure of a mutant EF-G reveals domain III and possibly the fusidic acid binding site. *J. Mol. Biol.* **303**, 593–603 (2000).
- Stark, H. *et al.* Visualization of elongation factor Tu on the *Escherichia coli* ribosome. *Nature* **389**, 403–406 (1997).
- Valle, M. *et al.* Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J.* **21**, 3557–3567 (2002).
- Stark, H. *et al.* Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. *Nature Struct. Biol.* **9**, 849–854 (2002).
- Agrawal, R. K., Penczek, P., Grassucci, R. A. & Frank, J. Visualization of elongation factor G on the *Escherichia coli* 70S ribosome: the mechanism of translocation. *Proc. Natl Acad. Sci. USA* **95**, 6134–6138 (1998).
- Stark, H., Rodnina, M. V., Wieden, H. J., van Heel, M. & Wintermeyer, W. Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation. *Cell* **100**, 301–309 (2000).
- Aevvarsson, A. *et al.* Three-dimensional structure of the ribosomal translocase: elongation factor G from *Thermus thermophilus*. *EMBO J.* **13**, 3669–3677 (1994).
- Czworkowski, J., Wang, J., Steitz, T. A. & Moore, P. B. The crystal structure of elongation factor G complexed with GDP, at 2.7 Å resolution. *EMBO J.* **13**, 3661–3668 (1994).
- Wilson, D. N. *et al.* in *The Ribosome. Structure, Function, Antibiotics and Cellular Interactions* (ed. Garrett, R. A.) 495–508 (American Society for Microbiology Press, Washington, DC, 2000).
- Aevvarsson, A. Structure-based sequence alignment of elongation factors Tu and G with related GTPases involved in translation. *J. Mol. Evol.* **41**, 1096–1104 (1995).
- Agrawal, R. K., Heagle, A. B., Penczek, P., Grassucci, R. A. & Frank, J. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nature Struct. Biol.* **6**, 643–647 (1999).
- Lancaster, L., Kiel, M. C., Kaji, A. & Noller, H. F. Orientation of ribosome recycling factor in the ribosome from directed hydroxyl radical probing. *Cell* **111**, 129–140 (2002).
- Mora, L., Zavialov, A., Ehrenberg, M. & Buckingham, R. H. Stop codon recognition and interactions with peptide release factor RF3 of truncated and chimeric RF1 and RF2 from *Escherichia coli*. *Mol. Microbiol.* **50**, 1467–1476 (2003).
- Zavialov, A. V., Mora, L., Buckingham, R. H. & Ehrenberg, M. Release of peptide promoted by the GGQ motif of class 1 release factors regulates the GTPase activity of RF3. *Mol. Cell* **10**, 789–798 (2002).
- van Heel, M. *et al.* Single-particle cryo electron microscopy: towards atomic resolution. *Quart. Rev. Biophys.* **33**, 307–369 (2000).
- van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
- Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **31**, 721–745 (2003).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Plewniak, F. *et al.* PipeAlign: a new toolkit for protein family analysis. *Nucleic Acids Res.* **31**, 3829–3832 (2003).
- Matadeen, R. *et al.* The *Escherichia coli* large ribosomal subunit at 7.5 Å resolution. *Structure Fold Des.* **15**, 1575–1583 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Ehrenberg and A. Zavialov for purified RF3 and 70S ribosome RCs; T. Pape for assistance in the early phase of this work; S. Chen, M. Schatz, R. Schmidt, and G. Willoughby for support; and R. Buckingham for sharing data before publication. B.P.K. thanks D. Moras, P. Schultz, M. Yusupov, G. Yusupova and B. Rees for support, discussions and comments on the manuscript; O. Poch for discussions on multiple sequence alignments; E. Orlova for initial image processing suggestions; and H. Saibil for the loan of a Tecnai 12 cryo-holder. The project was financed in part by grants from the Biotechnology and Biological Sciences Research Council, the European Union, and the Centre National de la Recherche Scientifique.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to B.P.K. (klaholz@igbmc.u-strasbg.fr). The cryo-EM maps of the RF3–70S complexes in states 1 and 2 have been deposited in the three-dimensional EM database (<http://www.ebi.ac.uk/msd/Services.html>) under accession codes EMD-1064 and EMD-1065, respectively.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequinot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Netherlands/

Italy/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatomachi,
Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Ghosts in the machine

These days, it is increasingly common to find that you can use the Internet to apply for a job or a grant. This might be convenient, but is it likely to be successful? For some who recently applied for research funding from the European Union (EU), the answer seems to be a resounding 'no'. After last month's closing date for the latest round of Marie Curie Fellowships, which fund postdocs to train in a foreign country, dozens of applicants complained about an apparent online glitch.

One French hopeful, who wants to remain anonymous in the hope that the EU will give her a second chance, spent weeks piecing together her online application, clicking the 'save' icon each time she updated it. Not seeing any obvious 'submit' icon on the web page, she thought that the final click on 'save' equated to submission. She waited to receive e-mail confirmation and, when she didn't, called Brussels, where she learned that several others had called with the same complaint.

The incident serves as a cautionary tale for a larger issue. Most pharmaceutical and large biotech companies feature similar human-resources websites that include online applications. Although they are designed for ease of use, the user-friendliness lies mostly on the surface. Applications through these recruitment sites are routinely scanned electronically for key words such as 'medicinal chemistry' or 'bioinformatics', the lack of which disqualifies the application just as easily as not clicking on 'submit'.

Anyone applying online for a grant or a job in any sector would do well to remember the importance of personal contact — people who hire and fund scientists are more reliable than machines. And, for the sake of the Marie Curie hopefuls whose applications got lost in the ether, perhaps the EU will also remember the importance of the human touch.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Getting mobile in Europe p868

CAREER VIEW

Scientists & Societies

The Polish biotech gap

Graduate Journal

One of those days

Movers

Louise Johnson

p870

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Young scientists studying abroad often hit roadblocks when heading home. Some of these are now being cleared away, says Sally

Last month, a group of expatriate postdocs sent an open letter of protest to the French government explaining why many of them would not be coming home. "Why return to France for a lower salary and deplorable working conditions?" they wrote.

Their action expressed the anger felt by many young European researchers who were encouraged to gain international experience, only to be let down when it came to coming back. But with competition for the best scientists becoming more voracious, research funding agencies in Europe are now making efforts



to ensure that the 'brain highway' is a two-way street.

For the past few decades, it has been largely one-way, to the United States. Now repatriation schemes run by the European Commission, programmes in individual countries and grants from foundations are helping researchers to return to their home continent. But before the flow of traffic back to Europe can be increased, some cultural and political roadblocks will need to be removed.

For Italian biochemist Davide Corona, that is already happening. Corona and his young family will be packing their bags in Santa Cruz next month and heading back to Italy after four years in the United States. Corona went to the University of California on a fellowship from the European Molecular Biology Organization, which was quickly exchanged for a more advantageous three-year Human Frontier Science Program (HFSP) fellowship. But after ten years living abroad, he is ready to go home.

"After I left Palermo, I thought I was in exile," says Corona. Academic tenure is hard enough to secure for researchers who stay in Italy and have 'sponsors'. Those who leave the country find that returning to a high-quality job is almost impossible.

HOMEWARD BOUND

But thanks in part to his HFSP fellowship, Corona secured a position at the San Raffaele Scientific Institute in Milan with funding from Telethon, an Italian charity that supports research into genetic diseases. Long-term HFSP fellowships let researchers put their third year of funding on hold for up to two years and, as Corona has done, use the money to help find a way back. The HFSP introduced this flexibility into its fellowships in 2000, recognizing the needs of internationally mobile researchers.

We want your brains

A number of European countries are funding award schemes to attract outstanding scientific talent from overseas. Although the number of these awards is small, they encourage senior researchers to leave the comfort of a well-established research group for new challenges.

Science Foundation Ireland (SFI), set up by the Irish government in 2000 with a remit to build up the country's neglected science base, has up to €1 million (US\$1.3 million) a year over five years to 'import' outstanding research talent. The SFI Research Professorship scheme has recently brought Irish-born molecular biologist John Atkins to the BioSciences Institute at University College Cork, where he will set up a new lab after nearly 20 years in the United States.

In the United Kingdom, the Royal Society's Wolfson Research Merit Awards provide £20 million (\$38 million) to help UK institutions attract or retain outstanding researchers. The awards, launched in 2001, boost salaries over five years and contribute to research costs. One beneficiary, John Schellnhuber, moved from the Potsdam Institute for Climate Impact Research in Germany, where he was founding director, to become research director of the Tyndall Centre for Climate Change Research in Norwich. Schellnhuber adds that the personal attention he received from the Royal Society and the UK government played a role. "It convinced me that there was top-level commitment for me to move to the United Kingdom and that I would be supported," he says. "That tipped the balance."

Schellnhuber is proud to have made the move at 52. "Most of us become mentally a bit sclerotic with age, repeating the same messages all the time," he says. "It is good to get 'shocked by displacement' in the later stages of our careers. I think it should be mandatory."

S.G.



The Researcher's Mobility Portal website provides country-by-country information about careers, grants and training opportunities. And 400 mobility centres opening in 30 EU and associated countries in May should help researchers to confront bureaucratic and cultural barriers. Each will provide face-to-face advice on matters such as administration and paperwork, childcare and housing.

HAMPERED BY THE SYSTEM

Nevertheless, the European Commission can only do so much to smooth the way. Many obstacles to mobility and integration are ingrained in national systems. Opaque recruitment practices are common in many countries, including Germany and Italy, while researchers in Spain tell Kafkaesque tales of their struggle to get foreign PhDs recognized. The frustration and discouragement among young researchers in many European labs make these unappealing destinations, despite the quality of their

Ease of passage: Raffaele Liberali (below left) hopes to make it easier for students such as Davide Corona (below) to return to their home countries after studying abroad.

work, to researchers who have tasted the freedom and resources available in many US labs.

After four years at the University of Washington in Seattle, computational biologist and HFSP fellow Tanja Kortemme accepted an assistant professorship at the University of California, San Francisco, rather than return home to Germany. "The trust that people put in you and the enormous resources available early in your career in the United States are tremendous," she says. "There just doesn't seem to be the same level of independence for young researchers in Europe."

The complexity of Europe's political infrastructure means changes will take time. The European scientific visa, approved this month, will cut non-European researchers' waiting time for a visa from four months to a few weeks, but is

unlikely to be fully in place until 2007. Many changes will require more commitment (and funding) from national governments.

Meanwhile the European Commission is looking at researchers' careers in their totality. A European Researchers' Charter for the management of research careers is due later this year, including guidelines on recognizing the value of mobility and a code of conduct for recruitment.

If the dream of a global jobs market is to become a reality, there needs to be a consistent effort by all players to create more opportunities for mobility, remove obstacles and harmonize career structures.

But until European governments take researcher protests seriously and make the necessary investment in change, there are likely to be more angry letters home from young scientists working overseas.

Sally Goodman is a freelance science writer based in Paris.



Other agencies are also becoming more flexible. The European Commission has introduced a range of new Marie Curie mobility grants. Its offer now includes repatriation grants and fellowships for researchers from outside the European Union (EU).

Raffaele Liberali, director of the commission's Human Factor, Mobility and Marie Curie Actions unit, has been fairly mobile himself recently, meeting groups of researchers in New York and Washington DC to raise awareness of opportunities for mobility in Europe. Many of the European postdocs asked him to reduce the period of research outside Europe needed to qualify for a Marie Curie repatriation grant. "They told me that five years was too long — after three years you either have tenure or you're out," says Liberali. "We will certainly be taking that into consideration."

Meanwhile, the unit has launched two new services.

T. GIPSTEIN/CORBIS

Web links

Human Frontier Science Program

• www.hfsp.org

Marie Curie Actions

• europa.eu.int/comm/research/fp6/mariecurie-actions/home_en.html

European Researcher's Mobility Portal and mobility centres

• europa.eu.int/eracareers

EMBO Life Sciences Mobility Portal

• mobility.embo.org
Science Foundation Ireland

• www.sfi.ie

Royal Society Wolfson Research Merit Awards

• www.royalsoc.ac.uk/funding/merit.htm

Early Stage Researcher Mobility in Europe conference

• www.mariecurie.org/esrm2004

GRADUATE JOURNAL

One of those days

Computers crashing for no apparent reason. Broken software. Code not compiling. Videoconferences. Phone conferences. Deadlines.

A PhD can be stressful. The science is challenging; we push ourselves to the limit. There are times when everything works and it's great. At other times, nothing seems to work — when I can't track down that annoying bug in my code, for example — and that is really frustrating. The key is not to give up. Be persistent. Sometimes I take a step back and remind myself of why I'm doing what I'm doing.

I've realized you must set your own standards. I'm my own boss; no one else is going to push me to get the job done. But this freedom can increase the stress level. I see many high-achievers around me who impose very high expectations on themselves. This can spill over onto others, making the environment both collaborative and competitive.

No one is immune to pressure, but it's important to maintain a sensible attitude. Why do I do this research? Because ultimately, it's rewarding. If it was easy I would get bored. As you progress, you learn how to stay sane and keep up with the pace. Well, I'm still sane and I'm still here, so I guess I'd better get back to that code. ■

Amber Jenkins is a graduate student in particle physics at Imperial College, London, doing thesis research at Fermilab in Batavia, Illinois.

SCIENTISTS & SOCIETIES

The Polish biotech gap

Like several other universities across the country, the Technical University of Lodz in Poland is actively training students to work in the biotechnology sector. But there is a problem: so far, Poland has very few biotech companies and there is scant venture capital available to start more.

So what should the graduate students do? They could bide their time doing a PhD in the hope that in the meantime things will improve, but funding is scarce, and even if they qualify, they would get only €180 (US\$232) per month. Poland's cost of living is lower than the European Union average, but it's still not low enough to live on that amount.

Rather than waiting and hoping, a group of professionals engaged in the life sciences are helping to develop the Polish Federation of

Biotechnology. We, as the young generation of biotechnologists, are looking forward to proposed initiatives such as establishing incubators and more academic–industry interaction.

We need to garner public support for biotechnology now, if incubators and funding are to follow later. To get that support, the Academic Students' Society of Biotechnology and the Young European Biotech Network hope to raise awareness through events in various European cities next year, starting in Lodz, Poland. The goal is to get people involved in a dialogue with life-sciences students and early-stage researchers.

Both politicians and the public must be made aware of how important biotechnology is to the future of Poland's economy. It is vital to show that it is the technology of the future and, consequently, that Poland will lose out if it

does not join the biotech race. If it continues to lag behind, it will end up importing all of the crucial technologies from abroad.

At present, biotechnology in Poland is a niche that is crying out for investment. The government is spending plenty of money on education, but if it were to invest in setting up biotech companies, it should readily see a return on its cash.

Hopefully, the situation will soon change. Once Poland becomes a full member of the European Union, more international companies will be interested in setting up laboratories in the country, and in investing in small biotech companies there. Graduate students will finally have the chance to develop their ideas and work in the field of their education. ■

Sylvia Gorlach and Mikolaj Slabicki are studying for an MSc in biotechnology at the Technical University of Lodz, Poland.
♦ <http://republika.pl/assb>

MOVERS

Louise Johnson, director of life sciences, Diamond Light Source, Chilton, UK



When Louise Johnson was a student at University College London, physics was confined to either the very large or the very small. She was hesitant to choose either, as they seemed to require huge teams to advance the field. "I didn't see how I could make a contribution," she says.

But then she recalled some X-ray crystallography work she had done as a

student at the Royal Institution in London. She was aware that the technique was catching on — the 1962 chemistry Nobel was awarded to Max Perutz and John Kendrew for their determination of the structure of the proteins myoglobin and haemoglobin using X-rays.

Three years later Johnson was part of a team that used the same technique to solve the structure of the enzyme lysozyme. Although Johnson wasn't on the main structure paper, she was co-author of a companion piece that showed how the enzyme's structure helped it to carry out its task of attacking bacterial invaders. "That's really what opened people's eyes to this method — that we could understand biology by understanding structure," she says.

From that moment on, Johnson knew that her career would be in structural biology. And that meant she needed to keep up with the latest technology to solve the structures of

bigger and more complicated proteins and complexes. Her first use of a synchrotron light source, LURE, near Paris, provided "a breath-taking moment" when she realized that its greater power could cut down a molecule's exposure to X-rays by two orders of magnitude. That development allowed her to help solve the structure of the enzyme glycogen phosphorylase.

Since then, she's pushed for stronger, more powerful facilities — and more facilities devoted to biology. In the early years, she jokes, researchers tackling biological problems were considered 'parasites' by scientists studying purely physical phenomena.

Now, the tables are turned. As life-sciences director of the Diamond synchrotron near Oxford, UK, Johnson's remit is to ensure that the new light source, when it comes online in 2007, devotes a significant amount of time to biology. ■

CV **1967–**: University of Oxford — demonstrator, zoology department (1967–73); lecturer in molecular biophysics (1973–90); professor in molecular biophysics (1990); professorial fellow Corpus Christi College (1990)
1967: Somerville College, Oxford — lecturer in biophysics (1967–73); additional fellow and lecturer (1973–90); honorary fellow (1990)
1966–67: Yale University, postdoc, biophysics department
1962–66: The Royal Institution, London (1965); PhD, University of London
1959–62: University College London, BSc physics